Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2219**

A DNA sequence (GBSx2338) was identified in *S.agalactiae* <SEQ ID 6847> which encodes the amino acid sequence <SEQ ID 6848>. Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF96286 GB:AE004374 hypothetical protein [Vibrio cholerae]
 Identities = 56/167 (33%), Positives = 89/167 (52%), Gaps = 12/167 (7%)

Query: 18  LAIIKSLPLNDCWLCAGTLRNFVWNKLS-GINETLTSDIDVVFFDKNI---SYEETVVLE 73
            L  +  L L  C++ AG +RN VW+ L   +  T  +DIDV+FFD +     YE++  LE
Sbjct: 41  LECVYQLELPQCYIAAGFVRNLVWDSLHHNVKLTPLNDIDVIFFDADCLDSDYEKS--LE 98

Query: 74  QQLKDNYPQYDWELKNEFYMNTHSPNTPKYTSSKDAISKFPEKCTAVGARLDDRNQLELY 133
             +L +  PQ +W++KN+  M+  + + P Y S+ DA+S +PEK TAV  R  + ++ E
Sbjct: 99  LKLSEQMPQLNWQVKNQAKMHLQNGDNP-YQSTLDAMSYWPEKETAVAVRKVEHDRYECI 157

Query: 134 LPYGEEEILNFIVSPTPYFEEDLLRYNVYLKRVDKKKWNNIWPRLTI 180
              +G E +   ++   P        Y ++  RV  K W  +WP L I
Sbjct: 158 SAFGFESLFQGFITHNP-----KRAYGIFENRVKSKGWLAMWPNLRI 199
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2220**

A DNA sequence (GBSx2339) was identified in *S.agalactiae* <SEQ ID 6849> which encodes the amino acid sequence <SEQ ID 6850>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2779(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB13060 GB:Z99110 yjdF [Bacillus subtilis]
 Identities = 47/138 (34%), Positives = 93/138 (67%), Gaps = 2/138 (1%)

Query: 1   MKMTVYFDGNFWLGLIEYDDDGDYKVFRYFFGKEPKDDDVFNFINHKLNDLIKKYEFVKT 60
            MK+T+Y+DG FW+G++E  D+G  + FR+ FGKEP+D +V  F++++L +++ + E  +
Sbjct: 24  MKLTIYYDGQFWVGVVEVVDNGKLRAFRHLFGKEPRDSEVLEFVHNQLLNMMAQAE--QE 81

Query: 61  DISLKRTNEHKKSPKRMQREINREKRKPVVSTKAQLAMKTIHMSIKNERQLSQKCKKNEL 120
             + L+   + K +PKR+QR++++E +   V++KAQ A+K    + K +++    K ++  +
Sbjct: 82  GVRLQGRRQKKINPKRLQRQVSKELKNAGVTSKAQEAIKLELEARKQKKKQIMKEQREHV 141
```

```
Query: 121 RKHRYQLKQEKRYQKKKG 138
           ++ RY LK++K  +K +G
Sbjct: 142 KEQRYMLKKQKAKKKHRG 159
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2221**

A DNA sequence (GBSx2340) was identified in *S.agalactiae* <SEQ ID 6851> which encodes the amino acid sequence <SEQ ID 6852>. This protein is predicted to be ComX1. Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3143(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9469> which encodes amino acid sequence <SEQ ID 9470> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD50429 GB:AF161701 ComX2 [Streptococcus pneumoniae]
 Identities = 61/152 (40%), Positives = 95/152 (62%)

Query: 5    EELFDKVKPIVMKLRRNYFVQLWEYDDWIQEGRIVLFRLLEEHPYLLDNESKLFIYFKTK 64
            +EL+++V+  V K R  Y++ LWE  DW QEG + L  L+     L+D+  +L  YFKTK
Sbjct: 3    KELYEEVQGTVYKCRNEYYLHLWELSDWDQEGMLCLHELISREEGLVDDIPRLRKYFKTK 62

Query: 65   FSNYLNDVLRHQDCQKRQFNKMPYEEISEVSHYVKSKGLVLDDYIAYRDTLTKVEETLSD 124
            F N + D +R Q+ QKR+++K PYEE+ E+SH +   GL LDDY  + +TL       S
Sbjct: 63   FRNRILDYIRKQESQKRRYDKEPYEEVGEISHRISEGGLWLDDYYLFHETLRDYRNKQSK 122

Query: 125 IDKEKFEKLISGERFAGKKQFIRDIQPFFNAF 156
              +E+ E+++S ERF G+++ +RD++  F  F
Sbjct: 123 EKQEELERVLSNERFRGRQRVLRDLRIVFKEF 154
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6853> which encodes the amino acid sequence <SEQ ID 6854>. Analysis of this protein sequence reveals the following:

```
     Possible site: 39

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-10.35    Transmembrane     9 -  25 (   7 -  28)

----- Final Results -----
               bacterial membrane --- Certainty=0.5140(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related sequence was also identified in GAS <SEQ ID 9163> which encodes the amino acid sequence <SEQ ID 9164>. Analysis of this protein sequence reveals the following:

```
     Possible site: 29
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-10.35    Transmembrane     2 -  18 (   1 -  18)
```

```
----- Final Results -----
               bacterial membrane --- Certainty=0.160(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAD50429 GB:AF161701 ComX2 [Streptococcus pneumoniae]
 Identities = 60/149 (40%), Positives = 98/149 (65%)

Query: 41  FEKVKPIILKLKRHYYIQLWDRDDWLQEGHIILLQLLERYPELIEEEERLYRYFKTKFSS 100
            +E+V+  + K +  YY+ LW+  DW QEG + L +L+ R   L+++  RL +YFKTKF +
Sbjct: 6   YEEVQGTVYKCRNEYYLHLWELSDWDQEGMLCLHELISREEGLVDDIPRLRKYFKTKFRN 65

Query: 101 YLKDLLRRQESQKRQFHKLAYEEIGEVAHAIPSRGLWLDDYVAYQEVIASLENQLNSQER 160
             + D +R+QESQKR++ K  YEE+GE++H I   GLWLDDY  + E +    N+ + +++
Sbjct: 66  RILDYIRKQESQKRRYDKEPYEEVGEISHRISEGGLWLDDYYLFHETLRDYRNKQSKEKQ 125

Query: 161 MQFQALIRGERFKGRRALLRKISPYFKEF 189
             + + ++  ERF+GR+ +LR +   FKEF
Sbjct: 126 EELERVLSNERFRGRQRVLRDLRIVFKEF 154
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 78/149 (52%), Positives = 116/149 (77%)

Query: 8   FDKVKPIVMKLRRNYFVQLWEYDDWIQEGRIVLFRLLEEHPYLLDNESKLFIYFKTKFSN 67
            F+KVKPI++KL+R+Y++QLW+ DDW+QEG I+L +LLE +P L++ E +L+ YFKTKFS+
Sbjct: 41  FEKVKPIILKLKRHYYIQLWDRDDWLQEGHIILLQLLERYPELIEEEERLYRYFKTKFSS 100

Query: 68  YLNDVLRHQDCQKRQFNKMPYEEISEVSHYVKSKGLVLDDYIAYRDTLTKVEETLSDIDK 127
            YL D+LR Q+ QKRQF+K+ YEEI EV+H + S+GL LDDY+AY++ +  +E  L+  ++
Sbjct: 101 YLKDLLRRQESQKRQFHKLAYEEIGEVAHAIPSRGLWLDDYVAYQEVIASLENQLNSQER 160

Query: 128 EKFEKLISGERFAGKKQFIRDIQPFFNAF 156
             +F+ LI GERF G++  +R I P+F  F
Sbjct: 161 MQFQALIRGERFKGRRALLRKISPYFKEF 189
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 2222**

A DNA sequence (GBSx2341) was identified in *S.agalactiae* <SEQ ID 6855> which encodes the amino acid sequence <SEQ ID 6856>. Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.23   Transmembrane  166 - 182 ( 166 - 182)

----- Final Results -----
               bacterial membrane --- Certainty=0.1893(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA99510 GB:Z75191 ORF YOR283w [Saccharomyces cerevisiae]
 Identities = 57/226 (25%), Positives = 97/226 (42%), Gaps = 22/226 (9%)

Query: 4   VRLYIARHGKTMFNTIGRAQGWSDTPLTTFGELGIKELGLGLKASNISFKEAFSSDSGRT 63
            +RL+I RHG+T  N     QG  DT +   GE    +LG  L++  I F +  SSD  R
Sbjct: 17  IRLFIIRHGQTEHNVKKILQGHKDTSINPTGEEQATKLGHYLRSRGIHFDKVVSSDLKRC 76
```

```
Query: 64  LQTMEIILREVQQENIPYTRDKRIREWCFGSLDGGYDGDLFNGVLPRVSNGDMSHLTHEE 123
            QT  ++L+  +QEN+P +    +RE   G ++G                  M     E+
Sbjct: 77  RQTTALVLKHSKQENVPTSYTSGLRERYMGVIEG------------------MQITEAEK 118

Query: 124 IANLICQVDTAGWAEPWAILSNRILSGFTAIAKKIEDIGGGNAIVVSHGMTIATFL-WL- 181
            A+   +     + E       R+         +  + G  N  +VSHG  I   L WL
Sbjct: 119 YADKHGEGSFRNFGEKSDDFVARLTGCVEEEVAEASNEGVKNLALVSHGGAIRMILQWLK 178

Query: 182 IDHSTPRSLGLDNGSVSVVDF--EDGTFSIQSIGDMSYREKGREIL 225
            ++     + + N SV++VD+  +   F ++ +G+  +   G  ++
Sbjct: 179 YENHQAHKIIVFNTSVTIVDYVKDSKQFIVRRVGNTQHLGDGEFVV 224
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6857> which encodes the amino acid sequence <SEQ ID 6858>. Analysis of this protein sequence reveals the following:

```
      Possible site: 57

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.69    Transmembrane  170 - 186 ( 170 - 186)

----- Final Results -----
               bacterial membrane --- Certainty=0.1277(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAA99510 GB:Z75191 ORF YOR283w [Saccharomyces cerevisiae]
 Identities = 64/231 (27%), Positives = 98/231 (41%), Gaps = 27/231 (11%)

Query: 5   RLYIARHGKTMFNTIGRAQGWSDTPLTKKGEEGIRELGLGLKDATIPFKAAFSSDSGRTM 64
           RL+I RHG+T  N     QG  DT +   GEE   +LG  L+   I F    SSD  R
Sbjct: 18  RLFIIRHGQTEHNVKKILQGHKDTSINPTGEEQATKLGHYLRSRGIHFDKVVSSDLKRCR 77

Query: 65  QTIEIILRESENEFLPYTKDNRIREWCFGSLEGTYDSELFLGVLPRTKAFENRDNLRDVP 124
           QT  ++L+ S+ E +P +  + +RE   G +EG   +E
Sbjct: 78  QTTALVLKHSKQENVPTSYTSGLRERYMGVIEGMQITEA--------------------- 116

Query: 125 YSELAESIVEVDTANWAEPWEVLRKRIWEGFEAIALSIQNAGGGNALVVSHGMTIGTFL- 183
             + A+   E    N+ E  +    R+    E       N G  N  +VSHG  I   L
Sbjct: 117 -EKYADKHGEGSFRNFGEKSDDFVARLTGCVEEEVAEASNEGVKNLALVSHGGAIRMILQ 175

Query: 184 WL--IDPDRDKQYIDNGSVTVVEF--DDGQFTIKTIGDMSYRYRGREIIEE 230
           WL   +    K  + N SVT+V++  D  QF ++ +G+  +   G  ++ +
Sbjct: 176 WLKYENHQAHKIIVFNTSVTIVDYVKDSKQFIVRRVGNTQHLGDGEFVVSD 226
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 150/231 (64%), Positives = 182/231 (77%), Gaps = 5/231 (2%)

Query: 1   MSKVRLYIARHGKTMFNTIGRAQGWSDTPLTTFGELGIKELGLGLKASNISFKEAFSSDS 60
           M+K RLYIARHGKTMFNTIGRAQGWSDTPLT  GE GI+ELGLGLK + I FK AFSSDS
Sbjct: 1   MTKTRLYIARHGKTMFNTIGRAQGWSDTPLTKKGEEGIRELGLGLKDATIPFKAAFSSDS 60

Query: 61  GRTLQTMEIILREVQQENIPYTRDKRIREWCFGSLDGGYDGDLFNGVLPRV----SNGDM 116
           GRT+QT+EIILRE + E +PYT+D RIREWCFGSL+G YD +LF GVLPR     +  ++
Sbjct: 61  GRTMQTIEIILRESENEFLPYTKDNRIREWCFGSLEGTYDSELFLGVLPRTKAFENRDNL 120

Query: 117 SHLTHEEIANLICQVDTAGWAEPWAILSNRILSGFTAIAKKIEDIGGGNAIVVSHGMTIA 176
             + + E+A  I +VDTA WAEPW +L  RI  GF AIA  I++ GGGNA+VVSHGMTI
Sbjct: 121 RDVPYSELAESIVEVDTANWAEPWEVLRKRIWEGFEAIALSIQNAGGGNALVVSHGMTIG 180

Query: 177 TFLWLIDHSTPRSLGLDNGSVSVVDFEDGTFSIQSIGDMSYREKGREILEK 227
           TFLWLID    +   +DNGSV+VV+F+DG F+I++IGDMSYR +GREI+E+
Sbjct: 181 TFLWLIDPDRDKQY-IDNGSVTVVEFDDGQFTIKTIGDMSYRYRGREIIEE 230
```

A related GBS gene <SEQ ID 8999> and protein <SEQ ID 9000> were also identified. Analysis of this protein sequence reveals the following:

```
Cytoplasmic predicted but experimentally found on the surface of Streptococci

32.3/52.0% over 184aa
                                        Thermotoga maritima
  EGAD|165681| phosphoglycerate mutase Insert characterized
  GP|4981935|gb|AAD36444.1|AE001791_6|AE001791 phosphoglycerate mutase Insert characterized
  PIR|G72260|G72260 phosphoglycerate mutase - (strain MSB8) Insert characterized

ORF01265(268 - 870 of 1248)
EGAD|165681|TM1374(1 - 185 of 201) phosphoglycerate mutase {Thermotoga maritima}
GP|4981935|gb|AAD36444.1|AE001791_6|AE001791 phosphoglycerate mutase {Thermotoga maritima}
PIR|G72260|G72260 phosphoglycerate mutase - Thermotoga maritima (strain MSB8)
%Match = 6.3
%Identity = 32.2  %Similarity = 52.0
Matches = 57  Mismatches = 78  Conservative Sub.s = 35

105       135       165       195       225       255       285       315
RGRNNSYEIFNPFSMLLKRINRFYFCSR*LQNFFIGKVR*YIPVKAFVFCYNIIKCL*GVSMSKVRLYIARHGKTMFNTI
                                                                ::||: |||:|::|
                                                                MKLYLIRHGETIWNEK
                                                                        10

345       375       405       435       465       495       519       549
GRAQGWSDTPLTTFGELGIKELGLGLKASNISFKEAFSSDSXRTLQTMEIILREVQQENI--PYTRDKRIREWCFGSLDG
|  || :| ||   |    ::|   ||         :||   |:|:| | |  ::| |     |:  |  |
GLWQGVTDVPLNERGREQARKLANSLK----RVDAIYSSPLKRSLETAEEIARRFEKEIIVEEDLRECEISLW-------
            30        40            50        60        70        80

579       609       639       669                 699       729       759
GYDGDLFNGVLPRVSNGDMSHLTHEEIANLICQVDTAGWA----------EPWAILSNRILSGFTAIAKKIEDIGGGNAI
                   : || || |     |:   |:          |    : ||::     |  : :  |  | :
-------------------NGLTVEE-AIREYPVEFKKWSSDPNFGMEGLESMRNVQNRVVKAIMKIVSQEKLNGSENVV
                      90        100       110       120       130       140

789       816       840       870       900       930       960       990
VVSHGMTIATFL-WLIDHST--PRSLGLDNGSVSVVDFEDGTFSIQSIGDMSYREKGREILEKTLQ*KKIKLSDSV*LVF
:||| :::  |: |::       |:: ||| |:|||: |     :        :|
IVSHSLSLRAFICWILGLPLYLHRNFKLDNASLSVVEIESKPRLVLLNDTCHLKES
            160       170       180       190       200
```

SEQ ID 9000 (GBS44) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 4 (lane 6; MW 27kDa), in Figure 168 (lane 8-10; MW 42kDa – thioredoxin fusion) and in Figure 238 (lane 7; MW 42kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 12 (lane 8; MW 52.4kDa).

Purified Thio-GBS44-His is shown in Figure 244, lanes 7 & 8.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## Example 2223

A DNA sequence (GBSx2342) was identified in *S.agalactiae* <SEQ ID 6859> which encodes the amino acid sequence <SEQ ID 6860>. This protein is predicted to be d-alanyl-d-alanine carboxypeptidase. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
```

```
                 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD00280 GB:U78599 putative D,D-carboxypeptidase [Streptococcus mutans]
 Identities = 108/169 (63%), Positives = 139/169 (81%)

Query: 79  ELSPDVVPVENIYLDKRITKQATQFLEAARAIDSREHLISGYRSVAYQEKLFNSYVTQEM 138
           E++PDV  ++ + +D RI +   +FL AA+ IDS EHLISGYRSVAYQE+L+N+Y+ QE
Sbjct: 4   EMNPDVTDIDGVKVDSRIAENTRKFLAAAQEIDSSEHLISGYRSVAYQEELYNNYIAQEK 63

Query: 139 TSNPNLTRGQAEKLVKTYSQPAGASEHQTGLAMDMSTVDSLNESDPRVVSQLKKIAPQYG 198
            +NP+L++ +A+K V+TYSQP G+SEHQTGLA+DMSTVDSLN+SD  VV+++  IAP+YG
Sbjct: 64  ANNPSLSQEEAQKQVQTYSQPPGSSEHQTGLAIDMSTVDSLNQSDANVVAKVAAIAPKYG 123

Query: 199 FVLRFPDGKTAETGVGYEDWHYRYVGVESAKYMAKHHLTLEEYITLLKE 247
           FVLRFP+GK   TG+ YEDWHYRYVGV+SAKYM KH LTLEEY+  LKE
Sbjct: 124 FVLRFPEGKKDATGIDYEDWHYRYVGVKSAKYMTKHDLTLEEYLKKLKE 172
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6861> which encodes the amino acid sequence <SEQ ID 6862>. Analysis of this protein sequence reveals the following:

```
      Possible site: 26
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -9.66   Transmembrane   10 -  26 (   3 -  29)

----- Final Results -----
                bacterial membrane --- Certainty=0.4864(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAD00280 GB:U78599 putative D,D-carboxypeptidase [Streptococcus mutans]
 Identities = 118/173 (68%), Positives = 139/173 (80%)

Query: 74  ITKEMSPELADINGISVDKRIEQATSDFLAAAQAIDLQEHLISGYRSVDYQTELYQSYIK 133
           IT EM+P++ DI+G+ VD RI + T  FLAAAQ ID  EHLISGYRSV YQ ELY +YI
Sbjct: 1   ITAEMNPDVTDIDGVKVDSRIAENTRKFLAAAQEIDSSEHLISGYRSVAYQEELYNNYIA 60

Query: 134 KEMANDPTLTQEAAEALVQTYSQPPGASEHHTGLAIDMSTVDTLNASDPSVAKAVQKIAP 193
           +E AN+P+L+QE A+  VQTYSQPPG+SEH TGLAIDMSTVD+LN SD +V   V  IAP
Sbjct: 61  QEKANNPSLSQEEAQKQVQTYSQPPGSSEHQTGLAIDMSTVDSLNQSDANVVAKVAAIAP 120

Query: 194 DYGFVLRFPEGKKTSTGVDYEDWHYRYVGKASARYMAQHNLTLEEYIAALKEK 246
            YGFVLRFPEGKK +TG+DYEDWHYRYVG  SA+YM +H+LTLEEY+  LKEK
Sbjct: 121 KYGFVLRFPEGKKDATGIDYEDWHYRYVGVKSAKYMTKHDLTLEEYLKKLKEK 173
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 131/235 (55%), Positives = 172/235 (72%), Gaps = 3/235 (1%)

Query: 15  LLAILCF--SLFALLKPNSQQSSSQKLRNEDIKKISSQKRNKKLQLPAVSSKDWNLILVN 72
           LL ++ F   L+  +KP    + +Q L  ++I++   +K ++   LP VS +DW L+LVN
Sbjct: 12  LLIVIVFLGGLYLFIKPEESVTPTQ-LNKKEIQQKDIKKTDRLRALPKVSVEDWELVLVN 70

Query: 73  RDHKHEELSPDVVPVENIYLDKRITKQATQFLEAARAIDSREHLISGYRSVAYQEKLFNS 132
           RDH  +E+SP++  +  I +DKRI +  + FL AA+AID +EHLISGYRSV YQ +L+ S
Sbjct: 71  RDHITKEMSPELADINGISVDKRIEQATSDFLAAAQAIDLQEHLISGYRSVDYQTELYQS 130

Query: 133 YVTQEMTSNPNLTRGQAEKLVKTYSQPAGASEHQTGLAMDMSTVDSLNESDPRVVSQLKK 192
           Y+ +EM ++P LT+  AE LV+TYSQP GASEH TGLA+DMSTVD+LN SDP V   ++K
Sbjct: 131 YIKKEMANDPTLTQEAAEALVQTYSQPPGASEHHTGLAIDMSTVDTLNASDPSVAKAVQK 190

Query: 193 IAPQYGFVLRFPDGKTAETGVGYEDWHYRYVGVESAKYMAKHHLTLEEYITLLKE 247
```

```
           IAP YGFVLRFP+GK   TGV YEDWHYRYVG  SA+YMA+H+LTLEEYI  LKE
Sbjct: 191 IAPDYGFVLRFPEGKKTSTGVDYEDWHYRYVGKASARYMAQHNLTLEEYIAALKE 245
```

A related GBS gene <SEQ ID 9001> and protein <SEQ ID 9002> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 7
McG: Discrim Score:     14.03
GvH: Signal Score (-7.5): -1.02
     Possible site: 27
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 0 value:  10.08 threshold:  0.0
   PERIPHERAL  Likelihood = 10.08      56
 modified ALOM score:  -2.52

*** Reasoning Step: 3

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
33.7/55.1% over 183aa

Enterococcus faecalis
  EGAD|41322| d-alanyl-d-alanine carboxypeptidase Insert characterized
GP|1209528|gb|AAB05624.1||U35369 D,D-carboxypeptidase Insert characterized

ORF01266(484 - 1038 of 1350)
EGAD|41322|43646(85  -  268  of  268)  d-alanyl-d-alanine  carboxypeptidase  {Enterococcus
faecalis}SP|Q47746|VANY_ENTFA D-ALANYL-D-ALANINE
 CARBOXYPEPTIDASE           (EC         3.4.16.4)           (DD-PEPTIDASE)            (DD-
CARBOXYPEPTIDASE).GP|1209528|gb|AAB05624.1||U35369 D,D-carboxypeptidase {Enteroco
ccus faecalis}
%Match = 10.1
%Identity = 33.7  %Similarity = 55.1
Matches = 63  Mismatches = 79  Conservative Sub.s = 40

234       264       294       324       354       384       414       444
SR*F*RWNIFYSIYWGYVLSRKRKRNFRKNIAMKKNKIIRFSLVGVLLAILCFSLFALLKPNSQQSSSQKLRNEDIKKIS

         MEKSNYHSNVNHHKRHMKQSGEKRAFLWAFIISFTVCTLFLGWRLVSVLEATQLPPIPATHTGSGTGVAEN
                 10        20        30        40        50        60        70

474       504       531       561       588       618       648       678
SQKRNKKLQLPAVSSKDWNLILVNRDHK-HEELSPDVVPVEN-IYLDKRITKQATQFLEAARAIDSREHLISGYRSVAYQ
 :             ::|:|||||| :    :   ::  : |   :| ||:     :::||||      : ||||:   |
PEENTLATAKEQGDEQEWSLILVNRQNPIPAQYDVELEQLSNGERIDIRISPYLQDLFDAARADGVYPIVASGYRTTEKQ
                 90       100       110       120       130       140       150

708       738       768       798       828       858       888       918
EKLFNSYVTQEMTSNPNLTRGQAEKLVKTYSQPAGASEHQTGLAMDMSTVDSLNESDPRVVSQLKKIAPQYGFVLRFPDG
::: :  |  |  :    |  ||:   :|:    | |||| |||:|::  | :: :   |   | : : ::||: |:|
QEIMDEKV-AEYKAK-GYTSAQAKAEAETWVAVPGTSEHQLGLAVDINA-DGIHSTGNEVYRWLDENSYRFGFIRRYPPD
       160        170       180       190        200       210       220

948       978       1008      1038      1068      1098      1128      1158
KTAETGVGYEDWHYRYVGVESAKYMAKHHLTLEEYITLLKENNQ*GNVFPC*ILLLLLLFSFSLFFFRF*TIREK*MLIL
||  |||  | |||||||:|:|  :  : | ||||:   |
KTEITGVSNEPWHYRYVGIEAATKIYHQGLCLEEYLNTEK
```

SEQ ID 6860 (GBS18) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 4 (lane 3; MW 31kDa).

The GBS18-His fusion product was purified (Figure 93A; see also Figure 189, lane 11) and used to immunise mice (lane 2 product; 20µg/mouse). The resulting antiserum was used for Western blot (Figure 93B), FACS (Figure 93C ), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

**Example 2224**

A DNA sequence (GBSx2343) was identified in *S.agalactiae* <SEQ ID 6863> which encodes the amino acid sequence <SEQ ID 6864>. This protein is predicted to be unnamed protein product. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-12.58    Transmembrane   10 -  26 (   3 -  29)

----- Final Results -----
              bacterial membrane --- Certainty=0.6031(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6865> which encodes the amino acid sequence <SEQ ID 6866>. Analysis of this protein sequence reveals the following:

```
     Possible site: 33
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-11.83    Transmembrane   10 -  26 (   4 -  33)

----- Final Results -----
              bacterial membrane --- Certainty=0.5734(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAD00279 GB:U78599 putative N-acetyl-muramidase [Streptococcus mutans]
 Identities = 66/150 (44%), Positives = 97/150 (64%), Gaps = 5/150 (3%)

Query: 18  LLLIVCPLLSSQRIASADKEVRVNYSQKQFITKMGKEVKPLAKYYGIRPSILIAQILLET 77
            LL+I+ P+L+S  +A A+K++   YS K+F+ ++    + L+K YG+R SI+I Q  L++
Sbjct: 3   LLVILLPILASGGLADANKKMPSPYSHKEFVKEIAPTAQKLSKIYGVRSSIIIGQAALDS 62

Query: 78  HDGKTLLASKYHNLFSKKATPGQVAITLKSPKQTN---QNV--RYAIYKDDASAIRDYLR 132
            H G TLLASKYHNLFS +A+PGQ A+ LKS +  N   Q V  RY +Y+    ++ DY+
Sbjct: 63  HFGSTLLASKYHNLFSIEASPGQGAVRLKSHEYKNGRWQEVTNRYLVYESWKESLYDYMA 122

Query: 133 MLRQGKEVDKRLYRNLATEKGYKAPAKSLQ 162
            +L   K  DK LY  + T  GYK  A++LQ
Sbjct: 123 ILHGNKIWDKALYTTMMTSSGYKTVARALQ 152
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 67/190 (35%), Positives = 102/190 (53%), Gaps = 1/190 (0%)

Query: 1   MRKRFSLLNFIVVTFIFFFFILFPLLNHKGKVDANSRQSVTYTKEEFIQKIVPDAQDLGK 60
            MRKR     F+ +   F   I+ PLL+ +    A+    V Y++++FI K+  + + L K
Sbjct: 1   MRKRLKFPYFLTLLACFLLLIVCPLLSSQRIASADKEVRVNYSQKQFITKMGKEVKPLAK 60

Query: 61  SYGIRPSFIIAQAALDSDFGEKILANKYHNLFGLLAEPGTPSITLNDSSTGKKQEKQFTH 120
             YGIRPS +IAQ  L++  G+ +LA+KYHNLF   A PG  +ITL  S     Q  ++
Sbjct: 61  YYGIRPSILIAQILLETHDGKTLLASKYHNLFSKKATPGQVAITLK-SPKQTNQNVRYAI 119

Query: 121 YKSWKYSMYDYLAHIKSGATGKKDSYTIMVSVKNPKTLVQKLQDSGFDNDKKYAKKMTEI 180
            YK    ++ DYL  ++ G    K  Y  + + K  K   + LQ      DK YA+++ ++
```

```
Sbjct: 120 YKDDASAIRDYLRMLRQGKEVDKRLYRNLATEKGYKAPAKSLQKYLHYTDKTYARRLIQV 179

Query: 181 IDLYDLTRYD 190
           I+  DLT YD
Sbjct: 180 IESNDLTNYD 189
```

SEQ ID 6864 (GBS246) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 61 (lane 7; MW 24.6kDa).

GBS246d was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 154 (lanes 14 & 15; MW 21kDa) and in Figure 183 (lane 4; MW 21kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 187 (lane 12; MW 46kDa). Purified GBS246d-GST is shown in Figure 243, lane 12.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## Example 2225

A DNA sequence (GBSx2344) was identified in *S.agalactiae* <SEQ ID 6867> which encodes the amino acid sequence <SEQ ID 6868>. Analysis of this protein sequence reveals the following:

```
Possible site: 44
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2541(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC45610 GB:U78296 repressor of class I heat shock gene
           expression HrcA [Streptococcus mutans]
 Identities = 227/345 (65%), Positives = 287/345 (82%), Gaps = 1/345 (0%)

Query: 17  VITQRQNDILNLIVELFTQTHEPVGSKALQRTIDSSSATIRNDMAKLEKLGLLEKAHTSS 76
           +ITQRQ DILNLIVELFT+THEP+GSK LQ +I SS ATIRNDMA LEKLGLLEKA T
Sbjct: 1   MITQRQKDILNLIVELFTKTHEPIGSKTLQNSIASSRATIRNDMAALEKLGLLEKATTPP 60

Query: 77  GRM-PSPAGFKYFVEHSLRLDSIDEQDIYHVIKAFDFEAFKLEDMLQKASHILSEMTGYT 135
             + P     +YFVEHSL  DS+DEQD+Y VIKAFDFEAF+L D+LQ+AS +L+ +TGYT
Sbjct: 61  AVVCPVKKAIRYFVEHSLNPDSLDEQDVYQVIKAFDFEAFRLGDLLQRASDVLANLTGYT 120

Query: 136 SVILDVEPARQRLTGFDVVQLSNHDALAVMTLDESKPVTVQFAIPRNFLTRDLIAFKAIV 195
           ++ILDVEP +QRLT FD+V+LSNHDALAV+TLDE+ PVTVQFAIP+NFL  DL+    I
Sbjct: 121 ALILDVEPKKQRLTTFDIVKLSNHDALAVLTLDEASPVTVQFAIPKNFLDSDLMTVAKIA 180

Query: 196 EERLLDGSVMDIHYKLRTEIPQIVQKYFVTTDNVLQLFDYVFSELFLETVFVAGKVNSLT 255
            ER L+ +V+DIHY+LRTE PQI+QKYF  TDNVL LFD++F+ +F E VF++GK+ +L
Sbjct: 181 RERFLNQTVLDIHYRLRTEPPQIIQKYFPRTDNVLDLFDHIFNPIFQEEVFISGKIKTLE 240

Query: 256 YSDLSTYQFLDNEQQVAISLRQSLKEGEMASVQVADSQEAALADVSVLTHKFLIPYRGFG 315
           ++ L TYQFL+N Q VA+ +RQSL E E+  VQVADS+E +LAD++V++ KFLIPYRGFG
Sbjct: 241 FAGLDTYQFLENLQSVALEIRQSLPEDELHRVQVADSKEKSLADLTVISQKFLIPYRGFG 300

Query: 316 LLSLIGPIDMDYRRSVSLVNIIGKVLAAKLGDYYRYLNSNHYEVH 360
           +L++IGP+D+DY+R++SL+N+I +VLA KLGD+YRYLNSNHYEVH
Sbjct: 301 ILTVIGPVDLDYQRTISLINVISRVLAVKLGDFYRYLNSNHYEVH 345
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6869> which encodes the amino acid sequence <SEQ ID 6870>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0695(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 341/344 (99%), Positives = 343/344 (99%)

Query: 17  VITQRQNDILNLIVELFTQTHEPVGSKALQRTIDSSSATIRNDMAKLEKLGLLEKAHTSS 76
            VITQRQNDILNLIVELFTQTHEPVGSKALQRTIDSSSATIRNDMAKLEKLGLLEKAHTSS
Sbjct: 1   VITQRQNDILNLIVELFTQTHEPVGSKALQRTIDSSSATIRNDMAKLEKLGLLEKAHTSS 60

Query: 77  GRMPSPAGFKYFVEHSLRLDSIDEQDIYHVIKAFDFEAFKLEDMLQKASHILSEMTGYTS 136
            GRMPSPAGFKYFVEHSLRLDSIDEQDIYHVIKAFDFEAFKLEDMLQKASHIL+EMTGYTS
Sbjct: 61  GRMPSPAGFKYFVEHSLRLDSIDEQDIYHVIKAFDFEAFKLEDMLQKASHILAEMTGYTS 120

Query: 137 VILDVEPARQRLTGFDVVQLSNHDALAVMTLDESKPVTVQFAIPRNFLTRDLIAFKAIVE 196
            VILDVEPARQRLTGFDVVQLSNHDALAVMTLDESKPVTVQFAIPRNFLTRDLIAFKAIVE
Sbjct: 121 VILDVEPARQRLTGFDVVQLSNHDALAVMTLDESKPVTVQFAIPRNFLTRDLIAFKAIVE 180

Query: 197 ERLLDGSVMDIHYKLRTEIPQIVQKYFVTTDNVLQLFDYVFSELFLETVFVAGKVNSLTY 256
            ERLLD SV+DIHYKLRTEIPQIVQKYFVTTDNVLQLFDYVFSELFLETVFVAGKVNSLTY
Sbjct: 181 ERLLDNSVIDIHYKLRTEIPQIVQKYFVTTDNVLQLFDYVFSELFLETVFVAGKVNSLTY 240

Query: 257 SDLSTYQFLDNEQQVAISLRQSLKEGEMASVQVADSQEAALADVSVLTHKFLIPYRGFGL 316
            SDLSTYQFLDNEQQVAISLRQSLKEGEMASVQVADSQEAALADVSVLTHKFLIPYRGFGL
Sbjct: 241 SDLSTYQFLDNEQQVAISLRQSLKEGEMASVQVADSQEAALADVSVLTHKFLIPYRGFGL 300

Query: 317 LSLIGPIDMDYRRSVSLVNIIGKVLAAKLGDYYRYLNSNHYEVH 360
            LSLIGPIDMDYRRSVSLVNIIGKVLAAKLGDYYRYLNSNHYEVH
Sbjct: 301 LSLIGPIDMDYRRSVSLVNIIGKVLAAKLGDYYRYLNSNHYEVH 344
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 2226**

A DNA sequence (GBSx2345) was identified in *S.agalactiae* <SEQ ID 6871> which encodes the amino acid sequence <SEQ ID 6872>. This protein is predicted to be grpe protein (grpE). Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5138(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC45611 GB:U78296 GrpE [Streptococcus mutans]
 Identities = 130/180 (72%), Positives = 151/180 (83%), Gaps = 3/180 (1%)

Query: 14  VSEEIKKDDLQEEVEATE--TEETVEEVIEEIPEKSELELANERADEFENKYLRAHAEM- 70
            +S++ KK++ +EEVEATE  TEE+VEEV EE  E  EL+ A ERA++FENKYLRAHAEM
Sbjct: 1   MSKKDKKEEYKEEVEATEPTTEESVEEVAEETSENKELQEALERAEDFENKYLRAHAEMP 60

Query: 71  QNIQRRSSEERQQLQRYRSQDLAKAILPSLDNLERALAVEGLTDDVKKGLEMTRDSLIQA 130
            +       +  +  QRYRSQDL KAILPSLDNLERALAVEGLTDDVKKGLEM ++SLIQA
Sbjct: 61  KTFSVALMKSDKVCQRYRSQDLRKAILPSLDNLERALAVEGLTDDVKKGLEMVQESLIQA 120
```

```
Query: 131 LKEEGVEEVEVDSFDHNFHMAVQTLPADDEHPADSIAEVFQKGYKLHERLLRPAMVVVYN 190
           LKEEGVEEVE+++FD N HMAVQTL ADD+HPADSIA+V QKGY+LHERLLRPAMVVVYN
Sbjct: 121 LKEEGVEEVELENFDANLHMAVQTLDADDDHPADSIAQVHQKGYQLHERLLRPAMVVVYN 180
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6873> which encodes the amino acid sequence <SEQ ID 6874>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5138(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 189/190 (99%), Positives = 189/190 (99%)

Query: 1   MAVFNKLFKRRHSVSEEIKKDDLQEEVEATETEETVEEVIEEIPEKSELELANERADEFE 60
           MAVFNKLFKRRHSVSEEIKKDDLQEEVEATETEETVEEVIEE PEKSELELANERADEFE
Sbjct: 1   MAVFNKLFKRRHSVSEEIKKDDLQEEVEATETEETVEEVIEETPEKSELELANERADEFE 60

Query: 61  NKYLRAHAEMQNIQRRSSEERQQLQRYRSQDLAKAILPSLDNLERALAVEGLTDDVKKGL 120
           NKYLRAHAEMQNIQRRSSEERQQLQRYRSQDLAKAILPSLDNLERALAVEGLTDDVKKGL
Sbjct: 61  NKYLRAHAEMQNIQRRSSEERQQLQRYRSQDLAKAILPSLDNLERALAVEGLTDDVKKGL 120

Query: 121 EMTRDSLIQALKEEGVEEVEVDSFDHNFHMAVQTLPADDEHPADSIAEVFQKGYKLHERL 180
           EMTRDSLIQALKEEGVEEVEVDSFDHNFHMAVQTLPADDEHPADSIAEVFQKGYKLHERL
Sbjct: 121 EMTRDSLIQALKEEGVEEVEVDSFDHNFHMAVQTLPADDEHPADSIAEVFQKGYKLHERL 180

Query: 181 LRPAMVVVYN 190
           LRPAMVVVYN
Sbjct: 181 LRPAMVVVYN 190
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 2227**

A DNA sequence (GBSx2346) was identified in *S.agalactiae* <SEQ ID 6875> which encodes the amino acid sequence <SEQ ID 6876>. This protein is predicted to be heat shock protein 70 (dnaK). Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0996(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6877> which encodes the amino acid sequence <SEQ ID 6878>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0996(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 594/609 (97%), Positives = 603/609 (98%), Gaps = 1/609 (0%)

Query: 1   MSKIIGIDLGTTNSAVAVLEGTESKIIANPEGNRTTPSVVSFKNGEIIVGDAAKRQAVTN 60
           MSKIIGIDLGTTNSAVAVLEGTESKIIANPEGNRTTPSVVSFKNGEIIVGDAAKRQAVTN
Sbjct: 1   MSKIIGIDLGTTNSAVAVLEGTESKIIANPEGNRTTPSVVSFKNGEIIVGDAAKRQAVTN 60

Query: 61  PDTVISIKSKMGTSEKVSANGKEYTPQEISAMILQYLKGYAEDYLGEKVEKAVITVPAYF 120
           P+TVISIKSKMGTSEKVSANGKEYTPQEISAMILQYLKGYAEDYLGEKVEKAVITVPAYF
Sbjct: 61  PETVISIKSKMGTSEKVSANGKEYTPQEISAMILQYLKGYAEDYLGEKVEKAVITVPAYF 120

Query: 121 NDAQRQATKDAGKIAGLEVERIVNEPTAAALAYGMDKTDKDEKILVFDLGGGTFDVSILE 180
           NDAQRQATKDAGKIAGLEVERIVNEPTAAALAYGMDKTDKDEKILVFDLGGGTFDVSILE
Sbjct: 121 NDAQRQATKDAGKIAGLEVERIVNEPTAAALAYGMDKTDKDEKILVFDLGGGTFDVSILE 180

Query: 181 LGDGVFDVLATAGDNKLGGDDFDQKIIDFLVEEFKKENGIDLSQDKMALQRLKDAAEKAK 240
           LGDGVFDVLATAGDNKLGGDDFDQKIIDFLV EFKKENGIDLSQDKMALQRLKDAAEKAK
Sbjct: 181 LGDGVFDVLATAGDNKLGGDDFDQKIIDFLVAEFKKENGIDLSQDKMALQRLKDAAEKAK 240

Query: 241 KDLSGVTQTQISLPFITAGSAGPLHLEMSLSRAKFDDLTRDLVERTKTPVRQALSDAGLS 300
           KDLSGVTQTQISLPFITAGSAGPLHLEMSLSRAKFDDLTRDLVERTKTPVRQALSDAGLS
Sbjct: 241 KDLSGVTQTQISLPFITAGSAGPLHLEMSLSRAKFDDLTRDLVERTKTPVRQALSDAGLS 300

Query: 301 LSEIDEVILVGGSTRIPAVVEAVKAETGKEPNKSVNPDEVVAMGAAIQGGVITGDVKDVV 360
           LSEIDEVILVGGSTRIPAVVEAVKAETGKEPNKSVNPDEVVAMGAAIQGGVITGDVKDVV
Sbjct: 301 LSEIDEVILVGGSTRIPAVVEAVKAETGKEPNKSVNPDEVVAMGAAIQGGVITGDVKDVV 360

Query: 361 LLDVTPLSLGIETMGGVFTKLIDRNTTIPTSKSQVFSTAADNQPAVDIHVLQGERPMAAD 420
           LLDVTPLSLGIETMGGVFTKLIDRNTTIPTSKSQVFSTAADNQPAVDIHVLQGERPMAAD
Sbjct: 361 LLDVTPLSLGIETMGGVFTKLIDRNTTIPTSKSQVFSTAADNQPAVDIHVLQGERPMAAD 420

Query: 421 NKTLGRFQLTDIPAAPRGIPQIEVTFDIDKNGIVSVKAKDLGTQKEQHIVIQSNSGLTDE 480
           NKTLGRFQLTDIPAAPRGIPQIEVTFDIDKNGIVSVKAKDLGTQKEQHIVI+SN GL++E
Sbjct: 421 NKTLGRFQLTDIPAAPRGIPQIEVTFDIDKNGIVSVKAKDLGTQKEQHIVIKSNDGLSEE 480

Query: 481 EIDKMMKDAEANAEADAKRKEEVDLKNEVDQAIFATEKTIKETEGKGFDTERDAAQSALD 540
           EID+MMKDAEANAEADAKRKEEVDLKNEVDQAIFATEKTIKETEGKGFDTERDAAQSALD
Sbjct: 481 EIDRMMKDAEANAEADAKRKEEVDLKNEVDQAIFATEKTIKETEGKGFDTERDAAQSALD 540

Query: 541 ELKKAQESGNLDDMKAKLEALNEKAQALAVKLYEQAAAAQQAAQGAEGAQSADSSSKGDD 600
           ELK AQESGNLDDMKAKLEALNEKAQALAVK+YEQAAAAQQAAQGAEGAQ+ DS++  DD
Sbjct: 541 ELKAAQESGNLDDMKAKLEALNEKAQALAVKMYEQAAAAQQAAQGAEGAQANDSAN-NDD 599

Query: 601 VVDGEFTEK 609
           VVDGEFTEK
Sbjct: 600 VVDGEFTEK 608
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## Example 2228

A DNA sequence (GBSx2347) was identified in *S.agalactiae* <SEQ ID 6879> which encodes the amino acid sequence <SEQ ID 6880>. This protein is predicted to be Streptococcus pneumoniae DnaJ protein homologue (dnaJ). Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4180(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6881> which encodes the amino acid sequence <SEQ ID 6882>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1322(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 330/377 (87%), Positives = 357/377 (94%), Gaps = 1/377 (0%)

Query: 1   MNNTEFYDRLGVSKDASQDEIKKAYRRMSKKYHPDINKETGAEEKYKEVQEAYETLSDTQ 60
           MNNTE+YDRLGVSKDASQD+IKKAYR+MSKKYHPDINKE GAE+KYK+VQEAYETLSD+Q
Sbjct: 19  MNNTEYYDRLGVSKDASQDDIKKAYRKMSKKYHPDINKEAGAEQKYKDVQEAYETLSDSQ 78

Query: 61  KRAAYDQYGAAGANGGFGGFDGGGFGGFDGGGFGGFEDIFSSFFGGGGMRNPNAPRQGDD 120
           KRAAYDQYGAAGA GGFGG   GGFGGFDGGGFGGFEDIFSSFFGGGG RNPNAPRQGDD
Sbjct: 79  KRAAYDQYGAAGAQGGFGG-GAGGFGGFDGGGFGGFEDIFSSFFGGGGSRNPNAPRQGDD 137

Query: 121 LQYRVNLSFEEAIFGAEKEVSYNRESSCHTCSGSGAKPGTSPVTCQKCHGSGVINVDTQT 180
           LQYRVNLSFEEA+FG EKEVSYNRE++C TC GSGAKPGT+PVTC+KCHGSGV+ +DTQT
Sbjct: 138 LQYRVNLSFEEAVFGVEKEVSYNREATCGTCLGSGAKPGTAPVTCRKCHGSGVMTIDTQT 197

Query: 181 PLGTMRRQVTCDVCQGSGQEIKEKCPTCHGTGHEKKTHKVSVKIPAGVETGQQIRLTGQG 240
           PLG MRRQVTCD+C GSG+EIKE C TCHGTGHEK+ HKVSVKIPAGVETGQQIRL GQG
Sbjct: 198 PLGMMRRQVTCDICHGSGKEIKEPCQTCHGTGHEKQAHKVSVKIPAGVETGQQIRLQGQG 257

Query: 241 EAGFNGGPYGDLFVIINVLPSQQFERNGSTIYYTLNISFVQAALGDTIDIPTVHGAVEMS 300
           EAGFNGGPYGDLFVI+NVLPS+QFERNGSTIYY L+ISF QAALGDT++IPTVHG VEM+
Sbjct: 258 EAGFNGGPYGDLFVILNVLPSKQFERNGSTIYYNLDISFTQAALGDTVEIPTVHGDVEMA 317

Query: 301 IPAGTQTGKTFRLRGKGAPKLRGGGQGDQHVTVNIVTPTKLNDAQKEALHAFAEASGDKM 360
           IPAGTQTGKTFRL+GKGAPKLRGGGQGDQHVTVNIVTPTKLNDAQ+EAL AFAEASG+KM
Sbjct: 318 IPAGTQTGKTFRLKGKGAPKLRGGGQGDQHVTVNIVTPTKLNDAQREALQAFAEASGEKM 377

Query: 361 VHPKKKGFFDKVKDALD 377
           +HPKKKGFFDKVKDAL+
Sbjct: 378 LHPKKKGFFDKVKDALE 394
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 2229**

A DNA sequence (GBSx2348) was identified in *S.agalactiae* <SEQ ID 6883> which encodes the amino acid sequence <SEQ ID 6884>. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.22   Transmembrane  281 - 297 ( 281 - 297)

----- Final Results -----
               bacterial membrane --- Certainty=0.1086(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD24445 GB:AF118389 unknown [Streptococcus suis]
 Identities = 182/373 (48%), Positives = 257/373 (68%), Gaps = 5/373 (1%)
```

```
Query: 4   KVEEIRSYLIASIQNGKLAPGDRLPSIRQLANQFSCNKDTVQRVLMELRFDNYIYAKPRS 63
           K + I   ++  I+  +   G++LPSIRQL  Q+ C+KDTVQ+ ++EL++ N IYA  +S
Sbjct: 3   KYQVIIQDILTGIEEHRFKRGEKLPSIRQLREQYHCSKDTVQKAMLELKYQNKIYAVEKS 62

Query: 64  GYYVFDSHQEEVEEGVSLPNSEIANIAYDDFRLCLNETLIGREDYLFNYYYRQEGLLDLS 123
           GYY+ +    + +      +  ++ I Y+DFR+CL E+LIGRE+YLFNYY++QEGL +L
Sbjct: 63  GYYILEDRDFQ-DHTCRAQSYRLSRITYEDFRICLKESLIGRENYLFNYYHQQEGLAELI 121

Query: 124 KAVAKLMEETGVYVPLDDIVITAGTQQALFILTQVTFPNRKSRVLIEEPTYPRMIELIKT 183
            +V  L+ +  VY   D +VITAG+QQAL+ILTQ+     K+ +LIE PTY RMIELI+
Sbjct: 122 SSVQSLLMDYHVYTKKDQLVITAGSQQALYILTQMETLAGKTEILIENPTYSRMIELIRH 181

Query: 184 QNLPYETISRGTHGIDFQRLEEIFQTQSIKFFYVIPRMHNPLGTSYNPVEMKRLIEMAEK 243
           Q +PY+TI R   GID + LE IFQT  IKFFY IPR+HNPLG++Y+      ++++A++
Sbjct: 182 QGIPYQTIERNLDGIDLEELESIFQTGKIKFFYTIPRLHNPLGSTYDIATKTAIVKLAKQ 241

Query: 244 YDVYIVEDDYMSDFASQS--PLHYYDTHGRVIYLKSFSKAIFPALRLAAICLPQALKSTF 301
           YDVYI+EDDY++DF S    PLHY DT  RVIY+KSF+  +FPALR+ AI LP  L+  F
Sbjct: 242 YDVYIIEDDYLADFDSSHSLPLHYLDTDNRVIYIKSFTPTLFPALRIGAISLPNQLRDIF 301

Query: 302 MAYKKLMDYDTNLILQKALALYIENGLYAKNSQYLKYRYQKDLANSKSILADHP-NLPSY 360
           + +K L+DYDTNLI+QKAL+LYI+NG++A+N+Q+L + Y       K  L  +  N+P Y
Sbjct: 302 IKHKSLIDYDTNLIMQKALSLYIDNGMFARNTQHLHHIYHAQWNKIKDCLEKYALNIP-Y 360

Query: 361 SLHHDSVLFDCSK 373
            +   SV F  SK
Sbjct: 361 RIPKGSVTFQLSK 373
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6885> which encodes the amino acid sequence <SEQ ID 6886>. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3043(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 176/382 (46%), Positives = 255/382 (66%), Gaps = 7/382 (1%)

Query: 1   MVTKVEEIRSYLIASIQNGKLAPGDRLPSIRQLANQFSCNKDTVQRVLMELRFDNYIYAK 60
           M TK + I S +   IQ  +L  GD+LPSIR L+  + C+KDTVQR L+EL++ + IYA
Sbjct: 1   MTTKYQTIISNIEQDIQKQRLKKGDKLPSIRVLSKVYYCSKDTVQRALLELKYRHLIYAV 60

Query: 61  PRSGYYVFDSHQEEVEEGVSLPNSEIANIAYDDFRLCLNETLIGREDYLFNYYYRQEGLL 120
           P+SGYYV   +    +  ++L   +  N+AY+DFRLCLNE L  ++ YLF+YY++ EGL
Sbjct: 61  PKSGYYVL-GNVSMPDNVLNLSLEDYNNMAYEDFRLCLNEALSAKDKYLFHYYHKTEGLE 119

Query: 121 DLSKAVAKLMEETGVYVPLDDIVITAGTQQALFILTQVTFPNRKSRVLIEEPTYPRMIEL 180
           +L +A+   + E  VY   D ++IT+GTQQAL+IL+Q+ FPN    +L+E+PTY RM  +
Sbjct: 120 ELREALLLYLAENSVYSNKDQLLITSGTQQALYILSQMPFPNTGKTILLEKPTYHRMEAI 179

Query: 181 IKTQNLPYETISRGTHGIDFQRLEEIFQTQSIKFFYVIPRMHNPLGTSYNPVEMKRLIEM 240
           +    LPY+TISR  +G+D + LE +FQT  IKFFY I R  +PLG SY+  E + ++ +
Sbjct: 180 VAQLGLPYQTISRHFNGLDLELLESLFQTGDIKFFYTISRFSHPLGLSYSTKEKEAIVRL 239

Query: 241 AEKYDVYIVEDDYMSDFA--SQSPLHYYDTHGRVIYLKSFSKAIFPALRLAAICLPQALK 298
           A++Y VYI+EDDY+ DF    + P+HYYDTH R+IYLKSFS ++FPALR+ A+ LP  LK
Sbjct: 240 AQRYQVYILEDDYLGDFVKLKEPPIHYYDTHHRIIYLKSFSMSVFPALRIGALVLPSGLK 299

Query: 299 STFMAYKKLMDYDTNLILQKALALYIENGLYAKNSQYLKYRYQKDLANSKSILADHPNLP 358
             F+  K L+D DTNL++QKALALY+ENG++ KN +++K RY K      ++     N P
Sbjct: 300 PHFLTQKSLIDLDTNLLMQKALALYLENGMFQKNLRFIK-RYLKQRERQLALFLKQ-NCP 357
```

```
Query: 359 S--YSLHHDSVLFDCSKLDNFK 378
              Y L    ++ D +  D+++
Sbjct: 358 DIHYQLTPTHLVIDYTTSDSYR 379
```

SEQ ID 6884 (GBS423) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 79 (lane 7; MW 49.3kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 172 (lane 2; MW 74kDa).

GBS423-GST was purified as shown in Figure 219, lane 2-3.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 2230**

A DNA sequence (GBSx2349) was identified in *S.agalactiae* <SEQ ID 6887> which encodes the amino acid sequence <SEQ ID 6888>. This protein is predicted to be pseudouridylate synthase I (truA). Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3265(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB03886 GB:AP001507 tRNA pseudouridine synthase A
          (pseudouridylate synthase I) [Bacillus halodurans]
 Identities = 105/240 (43%), Positives = 147/240 (60%), Gaps = 2/240 (0%)

Query: 1   MTRYKAQISYDGSAFSGFQRQPNCRTVQEEIERTLKRLNSGNDVIIHGAGRTDVGVHAYG 60
           M R   +++YDG+ F+G+Q QPN RTVQ E+E  LK ++ G  + +  +GRTD GVHA G
Sbjct: 1   MKRIGLKVAYDGTDFAGYQIQPNERTVQGELESVLKNIHKGMSIRVTASGRTDTGVHARG 60

Query: 61  QVIHFDLPQARDVEKLRFGLDTQCPDDIDIVKVEQVSDDFHCRYDKHIKTYEFLVDIGRP 120
           Q++HFD   +  V++    L++Q P DI +++   V  DFH RY    K Y + V
Sbjct: 61  QIVHFDTSLSFPVDRWPIALNSQLPADICVLEAADVPADFHARYSAKTKEYRYRVLTSAQ 120

Query: 121 KNPMMRNYATHYPYPVIIELMQEAIKDLVGTHDFTGFTASGTSVENKVRTIFDAKIQFEA 180
            +    RNY  H  YP+ +E MQ A   L+GTHDF+ F A+    VE+KVRTI D  + E
Sbjct: 121 ADVFRRNYTYHVRYPLDVEAMQRAAVQLLGTHDFSSFCAAKAEVEDKVRTIEDVALWREG 180

Query: 181 SKNLLIFTFTGNGFLYKQVRNMVGTLLKIGNGRMPISQIKTILQAKNRDLAGPTAAGNGL 240
            +  LIF+  GNGFLY  VR +VGTLL+IG G+    ++  IL A++R+ AG TA G+GL
Sbjct: 181 DE--LIFSIRGNGFLYNMVRIIVGTLLEIGAGKRSAEEVAKILAARSREAAGKTAPGHGL 238
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6889> which encodes the amino acid sequence <SEQ ID 6890>. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2558(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 184/249 (73%), Positives = 214/249 (85%)

Query: 1   MTRYKAQISYDGSAFSGFQRQPNCRTVQEEIERTLKRLNSGNDVIIHGAGRTDVGVHAYG 60
           M RYKA ISYDG+ FSGFQRQ + RTVQEEIE+TL +LN+G  +IIHGAGRTD GVHAYG
Sbjct: 1   MVRYKATISYDGTLFSGFQRQRHLRTVQEEIEKTLYKLNNGTKIIIHGAGRTDAGVHAYG 60

Query: 61  QVIHFDLPQARDVEKLRFGLDTQCPDDIDIVKVEQVSDDFHCRYDKHIKTYEFLVDIGRP 120
           QVIHFDLPQ ++VEKLRF LDTQ P+DID+V +E+V+DDFHCRY KH+KTYEFLVD GRP
Sbjct: 61  QVIHFDLPQEQEVEKLRFALDTQTPEDIDVVNIEKVADDFHCRYQKHLKTYEFLVDNGRP 120

Query: 121 KNPMMRNYATHYPYPVIIELMQEAIKDLVGTHDFTGFTASGTSVENKVRTIFDAKIQFEA 180
           KNPMMR+Y THYPY + I+LMQEAI  LVGTHDFTGFTA+GTSV+NKVRTI  A +  +
Sbjct: 121 KNPMMRHYTTHYPYTLNIKLMQEAINGLVGTHDFTGFTAAGTSVQNKVRTITKATVSRDE 180

Query: 181 SKNLLIFTFTGNGFLYKQVRNMVGTLLKIGNGRMPISQIKTILQAKNRDLAGPTAAGNGL 240
             + L+FTF+GNGFLYKQVRNMVGTLLKIGNG+MP+ Q+K IL +KNR LAGPT +GNGL
Sbjct: 181 KTDFLVFTFSGNGFLYKQVRNMVGTLLKIGNGQMPVEQVKVILSSKNRQLAGPTISGNGL 240

Query: 241 YLKEIIYED 249
           YLKEI YE+
Sbjct: 241 YLKEICYEN 249
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 2231**

A DNA sequence (GBSx2350) was identified in *S.agalactiae* <SEQ ID 6891> which encodes the amino acid sequence <SEQ ID 6892>. This protein is predicted to be phosphomethypyrimidine kinase (thiD). Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.2051(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB15828 GB:Z99123 phosphomethylpyrimidine kinase [Bacillus subtilis]
 Identities = 95/253 (37%), Positives = 150/253 (58%), Gaps = 13/253 (5%)

Query: 1   MKTRNVLAISGNDIFSGGGLHADLATYVVNKLHGFVAVTCLTAMSDKG---FEVIPIEAS 57
           M     L I+G+D   G G+ ADL T+    ++G  A+T + AM        +V PI+
Sbjct: 1   MSMHKALTIAGSDSSGGAGIQADLKTFQEKNVYGMTALTVIVAMDPNNSWNHQVFPIDTD 60

Query: 58  ILKQQLESLKD-VEFGSIKLGLLPNVETAQVVLEFVKSKQECPVVLDPVLVCKENHDL-- 114
            ++ QL ++ D +   ++K G+LP V+  ++  + +K KQ   VV+DPV+VCK  +++
Sbjct: 61  TIRAQLATITDGIGVDAMKTGMLPTVDIIELAAKTIKEKQLKNVVIDPVMVCKGANEVLY 120

Query: 115 --EVSQLREQLIAFFPYADVITPNLVEAQLLTGLS-IENLDQMKIAAEKLYDMGAKHVVI 171
                 LREQL    P A VITPNL EA  L+G+  ++ +D M  AA+K++ +GA++VVI
Sbjct: 121 PEHAQALREQLA---PLATVITPNLFEASQLSGMDELKTVDDMIEAAKKIHALGAQYVVI 177

Query: 172 KGGNRLNAEEATDLYYDGERFETYVFPVVDANNT-GAGCTFASSIASQLAMGKNVEDAVK 230
            GG +L  E+A D+ YDGE  E     ++D   T GAGCTF++++ ++LA G  V++A+
Sbjct: 178 TGGGKLKHEKAVDVLYDGETAEVLESEMIDTPYTHGAGCTFSAAVTAELAKGAEVKEAIY 237

Query: 231 MSKGFVYQAIKAS 243
            +K F+  AIK S
Sbjct: 238 AAKEFITAAIKES 250
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4407> which encodes the amino acid sequence <SEQ ID 4408>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2029(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 135/252 (53%), Positives = 174/252 (68%)

Query: 1   MKTRNVLAISGNDIFSGGGLHADLATYVVNKLHGFVAVTCLTAMSDKGFEVIPIEASILK 60
           MKT  ++ ISGNDI SGGGL+ADLATY+   L  FVAVTCLT  S++GF + P+   I +
Sbjct: 1   MKTDYIVTISGNDILSGGGLYADLATYIRYDLQAFVAVTCLTTRSEEGFSLFPVAKEIFR 60

Query: 61  QQLESLKDVEFGSIKLGLLPNVETAQVVLEFVKSKQECPVVLDPVLVCKENHDLEVSQLR 120
            QL S  +    +IK+GLLPN E  ++VL+F+K     PVVLDPVL CKE  D+++  LR
Sbjct: 61  DQLNSFTNAPISAIKIGLLPNAEMCEIVLDFIKGHLGIPVVLDPVLACKEIDDVKIVPLR 120

Query: 121 EQLIAFFPYADVITPNLVEAQLLTGLSIENLDQMKIAAEKLYDMGAKHVVIKGGNRLNAE 180
           ++++   PY  V+TPNLVEAQLL+   I +L  M+ AA+  Y +GAK VVIKGGNR + +
Sbjct: 121 QEILQLLPYVTVVTPNLVEAQLLSQKEIVSLKDMQEAAKYFYQLGAKQVVIKGGNRFSQK 180

Query: 181 EATDLYYDGERFETYVFPVVDANNTGAGCTFASSIASQLAMGKNVEDAVKMSKGFVYQAI 240
           +A DL+YDG+   T   PV++ NN GAGCTFASSIASQL   K   +AVK SK  VYQAI
Sbjct: 181 KAIDLFYDGKEIVTLECPVLEKNNIGAGCTFASSIASQLVKKKTPLEAVKNSKELVYQAI 240

Query: 241 KASDKYGVVQHF 252
             SD+YGV Q +
Sbjct: 241 LQSDRYGVKQSY 252
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 2232**

A DNA sequence (GBSx2351) was identified in *S.agalactiae* <SEQ ID 6893> which encodes the amino acid sequence <SEQ ID 6894>. Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -6.05   Transmembrane   97 - 113 (  96 - 119)
   INTEGRAL    Likelihood = -0.22   Transmembrane   54 -  70 (  54 -  70)

----- Final Results -----
               bacterial membrane --- Certainty=0.3421(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAA30952 GB:AP000007 202aa long hypothetical protein [Pyrococcus
           horikoshii]
 Identities = 48/148 (32%), Positives = 78/148 (52%), Gaps = 9/148 (6%)

Query: 10  VQLAIVTAISIVLGMFISIPTPTGFLTLLDAGIFFAAFYFGKKEGAVVGALAGFLIDLLK 69
           V  A+VTA+++V+   I IP   G+L   D  I   +  FG   G   G +     DLL
Sbjct: 49  VMAALVTAMTMVIR--IPIPASQGYLNFGDIMIMLTSVLFGPLVGGFAGGVGSAFADLL- 105

Query: 70  GYPNWMFFSLLIHGTQGYLAGLPGR------RRLLGLISATLVMVLGYAIASGLMYGWGA 123
           GYP+W  F+L+I GT+G + G   +      + LLG +    VMV+GY   + ++YG
```

```
Sbjct: 106 GYPSWALFTLVIKGTEGIIVGYFSKGEANYGKILLGTVLGGSVMVIGYVSVAYVLYGPAG 165

Query: 124 VLPDIPGNIMQNMVGMVVGFALNKSLER 151
            + ++  +I+Q + G+V+G  L   L++
Sbjct: 166 AIGELYNDIVQAVSGIVIGGGLGYILKK 193
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6895> which encodes the amino acid sequence <SEQ ID 6896>. Analysis of this protein sequence reveals the following:

```
     Possible site: 54

>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -4.62   Transmembrane   98 - 114 (  97 - 119)
   INTEGRAL    Likelihood = -0.00   Transmembrane  135 - 151 ( 135 - 151)

----- Final Results -----
               bacterial membrane --- Certainty=0.2848(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB49310 GB:AJ248284 hypothetical protein [Pyrococcus abyssi]
 Identities = 42/145 (28%), Positives = 73/145 (49%), Gaps = 10/145 (6%)

Query: 7   RQMSLTGILTALVVVLGRFVMLPTPT--GFLTLLDAGIYAVSFSFGSAQGAIVGGLSGFL 64
           R ++++ +  ALV  +   + +P P   G+L   D  I  V+  FG   G   GG+   +
Sbjct: 39  RTVAISAVAAALVTAMTMVIRIPIPASQGYLNFGDIMIMLVAVLFGPLVGGFAGGVGSAI 98

Query: 65  IDLVAGYPQWMFHSLIAHSVQGYFAGWRGR------KRWLGVVIGSFIMIFWYFLGSLML 118
            DL+ GYP W   +LI    +G   G+  +      K  +G V+G FIM+  Y   S +L
Sbjct: 99  ADLI-GYPSWALFTLIIKGSEGLVVGYFSKGEPNYSKILIGTVLGGFIMVLGYVSVSYVL 157

Query: 119 GYGLSGSLAGIWGNVMQNTLGLFVG 143
            YG +G+++ ++ + +Q   G+ +G
Sbjct: 158 -YGPAGAISELYNDTVQAVSGIVIG 181
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 77/155 (49%), Positives = 106/155 (67%), Gaps = 1/155 (0%)

Query: 1   MRKEKTSQLVQLAIVTAISIVLGMFISIPTPTGFLTLLDAGIFFAAFYFGKKEGAVVGAL 60
           M+  K  Q+    I+TA+ +VLG F+ +PTPTGFLTLLDAGI+  +F FG  +GA+VG L
Sbjct: 1   MQNSKIRQMSLTGILTALVVVLGRFVMLPTPTGFLTLLDAGIYAVSFSFGSAQGAIVGGL 60

Query: 61  AGFLIDLLKGYPNWMFFSLLIHGTQGYLAGLPGRRRLLGLISATLVMVLGYAIASGLM-Y 119
           +GFLIDL+ GYP WMF SL+ H  QGY AG  GR+R LG++  + +M+  Y + S ++ Y
Sbjct: 61  SGFLIDLVAGYPQWMFHSLIAHSVQGYFAGWRGRKRWLGVVIGSFIMIFWYFLGSLMLGY 120

Query: 120 GWGAVLPDIPGNIMQNMVGMVVGFALNKSLERVKK 154
           G    L  I GN+MQN +G+ VGF + K++ R KK
Sbjct: 121 GLSGSLAGIWGNVMQNTLGLFVGFIIFKAILRQKK 155
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 2233**

A DNA sequence (GBSx2352) was identified in *S.agalactiae* <SEQ ID 6897> which encodes the amino acid sequence <SEQ ID 6898>. Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
```

```
                bacterial cytoplasm --- Certainty=0.0881(Affirmative) < succ>
                 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB15708 GB:Z99122 alternate gene name: ipc-33d [Bacillus subtilis]
 Identities = 91/176 (51%), Positives = 115/176 (64%)

Query: 6   NKLKQETKAIVVDIIERSALKKGQIFVLGLSSSEVSGGLIGKNSSSEIGEIIVEVILKEL 65
            N+LKQ  K ++ +  +++ LK+ Q+FVLG S+SEV+G  IG + S +I E I   + +
Sbjct: 2   NELKQTWKTMLSEFQDQAELKQDQLFVLGCSTSEVAGSRIGTSGSVDIAESIYSGLAELR 61

Query: 66  HSRGIYLAVQGCEHVNRALVVEAELAERQQLEVVNVVPNLHAGGSGQVAAFKLMTSPVEV 125
               GI+LA Q CEH+NRALVVEAE A+  +L  V+ VP   AGG+    AFK M SPV V
Sbjct: 62  EKTGIHLAFQCCEHLNRALVVEAETAKLFRLPTVSAVPVPKAGGAMASYAFKQMKSPVLV 121

Query: 126 EEIVAHAGIDIGDTSIGMHIKRVQVPLIPISRELGGAHVTALASRPKLIGGARAGY 181
            E I A AGIDIGDT IGMH+K V VP+      LG AHVT   +RPKLIGG RA Y
Sbjct: 122 ETIQADAGIDIGDTFIGMHLKPVAVPVRVSQNSLGSAHVTLARTRPKLIGGVRAVY 177
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6899> which encodes the amino acid sequence <SEQ ID 6900>. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2166(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 132/183 (72%), Positives = 161/183 (87%)

Query: 6   NKLKQETKAIVVDIIERSALKKGQIFVLGLSSSEVSGGLIGKNSSSEIGEIIVEVILKEL 65
            N L+++T+ IV+D++ERSA++ G +FVLGLSSSE+ G  IGK SS E+G+I+VEV+L EL
Sbjct: 3   NNLEKQTREIVIDVVERSAIQPGNLFVLGLSSSEILGSRIGKQSSLEVGQIVVEVVLDEL 62

Query: 66  HSRGIYLAVQGCEHVNRALVVEAELAERQQLEVVNVVPNLHAGGSGQVAAFKLMTSPVEV 125
            + RG++LAVQGCEHVNRALVVE  +AE +QLE+VNVVPNLHAGGS Q+AAF+LM+ PVEV
Sbjct: 63  NKRGVHLAVQGCEHVNRALVVERHVAESKQLEIVNVVPNLHAGGSAQMAAFQLMSDPVEV 122

Query: 126 EEIVAHAGIDIGDTSIGMHIKRVQVPLIPISRELGGAHVTALASRPKLIGGARAGYTSDP 185
            EE++AHAG+DIGDT+IGMHIKRVQ+PLIP  RELGGAHVTALASRPKLIGGARA Y  D
Sbjct: 123 EEVIAHAGLDIGDTAIGMHIKRVQIPLIPCQRELGGAHVTALASRPKLIGGARADYNMDI 182

Query: 186 IRK 188
            IRK
Sbjct: 183 IRK 185
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 2234**

A DNA sequence (GBSx2353) was identified in *S.agalactiae* <SEQ ID 6901> which encodes the amino acid sequence <SEQ ID 6902>. Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-11.25    Transmembrane    21 -  37 (  13 -  46)
   INTEGRAL    Likelihood = -4.30    Transmembrane    78 -  94 (  76 - 113)
   INTEGRAL    Likelihood = -2.07    Transmembrane    96 - 112 (  95 - 113)
```

```
----- Final Results -----
               bacterial membrane --- Certainty=0.5501(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB06385 GB:AP001516 unknown conserved protein [Bacillus halodurans]
 Identities = 105/261 (40%), Positives = 150/261 (57%), Gaps = 2/261 (0%)

Query: 12  NVEEVLFTFFTKLIS--ILLLIIAFVIVRQVINYLFEKTVNRSLAFSRQKVARQKTLAKL 69
            N+    F   T +I+  +L+ +IAF+IVR +   +   +  R    ++    R  TL KL
Sbjct: 7   NITSGAFLASTFIIAGKVLVAVIAFLIVRAIGKRIISNSFARMAKNNQLSSGRVVTLEKL 66

Query: 70  SHNVLNYTLYFFLFYWILSILGVPISSLLAGAGIAGVAIGLGAQGFLSDVVNGFFILLEN 129
            S N  +YTL F     +L+I G+  S+L+AGAGI G+AIG GAQG +SD+V GFFILLE
Sbjct: 67  SLNAFSYTLMFIFATTLLTIFGLNPSALIAGAGIVGLAIGFGAQGLVSDIVTGFFILLEK 126

Query: 130 QFDVGDIINVGTVSGTVTNVGIRTTQIHDFDGTLHFIPNRNITIVSNKSRSNMRAQIDIP 189
            Q DVGD +  G V G V  VG+RT  I  FDGTLH+IPNRNI  VSN SR NMRA +DI
Sbjct: 127 QIDVGDYVTAGGVDGIVEEVGLRTALIRGFDGTLHYIPNRNIANVSNHSRGNMRALVDIS 186

Query: 190 LFVHTNLDQISDIVTKINEEYVSKHPAIVGEPTVFGPTTNANGQFVYRINIFTQNGAQFD 249
            +  + N+D+   ++ K+ ++   +   I+  P V G     +   V RI   T+N  Q+
Sbjct: 187 ISYNDNIDEAISVMQKVCDQLAEQDERIIEGPDVIGVQNLGDSDVVIRIIAKTENMEQWS 246

Query: 250 IYAEFYKLYQKAILEEGIDLP 270
            +     K  ++A+    I++P
Sbjct: 247 VERLLRKQLKEALEAHNIEIP 267
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6903> which encodes the amino acid sequence <SEQ ID 6904>. Analysis of this protein sequence reveals the following:

```
     Possible site: 54
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.49   Transmembrane   24 -  40 (  15 -  45)
   INTEGRAL    Likelihood = -4.83   Transmembrane   78 -  94 (  73 -  99)
   INTEGRAL    Likelihood = -2.07   Transmembrane   96 - 112 (  95 - 113)

----- Final Results -----
               bacterial membrane --- Certainty=0.4397(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:BAB06385 GB:AP001516 unknown conserved protein [Bacillus halodurans]
 Identities = 104/249 (41%), Positives = 151/249 (59%), Gaps = 4/249 (1%)

Query: 22  KKLVSLIILLLFFAILKRVTNYLFEKTINKSFAYSRQSEARKKTLSKLTHNILNYLLYFL 81
            K LV++I  L+  AI KR+ +  F +    +    + S  R  TL KL+ N  +Y L F+
Sbjct: 23  KVLVAVIAFLIVRAIGKRIISNSFARMAKNN----QLSSGRVVTLEKLSLNAFSYTLMFI 78

Query: 82  LIYWILSLFGIPVSSLLAGAGIAGVAIGLGAQGFLSDVVNGFFILFENQFEVGDNVTISD 141
                +L++FG+  S+L+AGAGI G+AIG GAQG +SD+V GFFIL E Q +VGD VT
Sbjct: 79  FATTLLTIFGLNPSALIAGAGIVGLAIGFGAQGLVSDIVTGFFILLEKQIDVGDYVTAGG 138

Query: 142 IEGSVFGVGIRTTQIRGFDGTLHFIPNRSITVVSNKSRGNMRALIEIPLYSTVNLSQVTR 201
            ++G V  VG+RT  IRGFDGTLH+IPNR+I  VSN SRGNMRAL++I +    N+ +
Sbjct: 139 VDGIVEEVGLRTALIRGFDGTLHYIPNRNIANVSNHSRGNMRALVDISISYNDNIDEAIS 198

Query: 202 IIDEVNQKELPNHPQIVGKPNILGPQNNSNGQFTFRIAIFTENGEQFKIYHTFYRLYQEA 261
            ++ +V  +      +I+  P+++G QN  +     RI   TEN EQ+ +     +  +EA
Sbjct: 199 VMQKVCDQLAEQDERIIEGPDVIGVQNLGDSDVVIRIIAKTENMEQWSVERLLRKQLKEA 258

Query: 262 LLKEGIQLP 270
```

```
           L    I++P
Sbjct: 259 LEAHNIEIP 267
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 164/265 (61%), Positives = 215/265 (80%)

Query: 7   FIDHLNVEEVLFTFFTKLISILLLIIAFVIVRQVINYLFEKTVNRSLAFSRQKVARQKTL 66
           +++  ++E +  T F KL+S+++L++ F I+++V NYLFEKT+N+S A+SRQ  AR+KTL
Sbjct: 7   YLEQSHIENIGLTIFKKLVSLIILLLFFAILKRVTNYLFEKTINKSFAYSRQSEARKKTL 66

Query: 67  AKLSHNVLNYTLYFFLFYWILSILGVPISSLLAGAGIAGVAIGLGAQGFLSDVVNGFFIL 126
           +KL+HN+LNY LYF L YWILS+ G+P+SSLLAGAGIAGVAIGLGAQGFLSDVVNGFFIL
Sbjct: 67  SKLTHNILNYLLYFLLIYWILSLFGIPVSSLLAGAGIAGVAIGLGAQGFLSDVVNGFFIL 126

Query: 127 LENQFDVGDIINVGTVSGTVTNVGIRTTQIHDFDGTLHFIPNRNITIVSNKSRSNMRAQI 186
            ENQF+VGD + +  + G+V  VGIRTTQI  FDGTLHFIPNR+IT+VSNKSR NMRA I
Sbjct: 127 FENQFEVGDNVTISDIEGSVFGVGIRTTQIRGFDGTLHFIPNRSITVVSNKSRGNMRALI 186

Query: 187 DIPLFVHTNLDQISDIVTKINEEYVSKHPAIVGEPTVFGPTTNANGQFVYRINIFTQNGA 246
           +IPL+   NL Q++ I+ ++N++ +  HP IVG+P + GP  N+NGQF +RI IFT+NG
Sbjct: 187 EIPLYSTVNLSQVTRIIDEVNQKELPNHPQIVGKPNILGPQNNSNGQFTFRIAIFTENGE 246

Query: 247 QFDIYAEFYKLYQKAILEEGIDLPT 271
           QF IY  FY+LYQ+A+L+EGI LPT
Sbjct: 247 QFKIYHTFYRLYQEALLKEGIQLPT 271
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 2235**

A DNA sequence (GBSx2354) was identified in *S.agalactiae* <SEQ ID 6905> which encodes the amino acid sequence <SEQ ID 6906>. This protein is predicted to be RopA (tig). Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1785(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9283> which encodes amino acid sequence <SEQ ID 9284> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6907> which encodes the amino acid sequence <SEQ ID 6908>. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0776(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 303/354 (85%), Positives = 337/354 (94%)

Query: 1   MSTSFENKATNRGIITFTISQDEIKPALDQAFNKVKKDLNVPGFRKGHMPRTVFNQKFGE 60
```

```
           MSTSFENKATNRG+ITFTISQD+IKPALD+AFNK+KKDLN PGFRKGHMPR VFNQKFGE
Sbjct: 30  MSTSFENKATNRGVITFTISQDKIKPALDKAFNKIKKDLNAPGFRKGHMPRPVFNQKFGE 89

Query: 61  EALYENALNLVLPKAYEAAVAELGLDVVAQPKIDVVSMEKGQDWKLTAEVVTKPEVKLGD 120
           E LYE+ALN+VLP+AYEAAV ELGLDVVAQPKIDVVSMEKG++W L+AEVVTKPEVKLGD
Sbjct: 90  EVLYEDALNIVLPEAYEAAVTELGLDVVAQPKIDVVSMEKGKEWTLSAEVVTKPEVKLGD 149

Query: 121 YKDLSVEVDASKEVSDEEVDAKVERERNNLAELTVKDGEAAQGDTVVIDFVGSVDGVEFD 180
           YK+L VEVDASKEVSDE+VDAK+ERER NLAEL +KDGEAAQGDTVVIDFVGSVDGVEFD
Sbjct: 150 YKNLVVEVDASKEVSDEDVDAKIERERQNLAELIIKDGEAAQGDTVVIDFVGSVDGVEFD 209

Query: 181 GGKGDNFSLELGSGQFIPGFEEQLVGSKAGQTVDVNVTFPEDYQAEDLAGKDAKFVTTIH 240
           GGKGDNFSLELGSGQFIPGFE+QLVG+KAG  V+VNVTFPE YQAEDLAGK AKF+TTIH
Sbjct: 210 GGKGDNFSLELGSGQFIPGFEDQLVGAKAGDEVEVNVTFPESYQAEDLAGKAAKFMTTIH 269

Query: 241 EVKTKEVPALDDELAKDIDDEVETLDELKAKYRKELESAKEIAFDDAVEGAAIELAVANA 300
           EVKTKEVP LDDELAKDID++V+TL++LK KYRKELE+A+E A+DDAVEGAAIELAVANA
Sbjct: 270 EVKTKEVPELDDELAKDIDEDVDTLEDLKVKYRKELEAAQETAYDDAVEGAAIELAVANA 329

Query: 301 EIVELPEEMVHDEVHRAMNEFMGNMQRQGISPEMYFQLTGTTEEDLHKQYQADA 354
           EIV+LPEEM+H+EV+R++NEFMGNMQRQGISPEMYFQLTGTT+EDLH QY A+A
Sbjct: 330 EIVDLPEEMIHEEVNRSVNEFMGNMQRQGISPEMYFQLTGTTQEDLHNQYSAEA 383
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 2236**

A DNA sequence (GBSx2355) was identified in *S.agalactiae* <SEQ ID 6909> which encodes the amino acid sequence <SEQ ID 6910>. This protein is predicted to be galactose-6-phosphate isomerase laca subunit (rpiB). Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3491(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAA25177 GB:M60447 galactose 6-P isomerase [Lactococcus lactis]
 Identities = 92/141 (65%), Positives = 115/141 (81%)

Query: 1   MTIIIGADAHGVELKEVIRQHLTSLGKEIIDLTDTSKDFVDNTLAIVAKVNQKEDNLGIM 60
           M I++GAD  G  LK+V++  L   G E+ID+T   +DFVD TLA+ ++VN+ E NLGI+
Sbjct: 1   MAIVVGADLKGTRLKDVVKNFLVEEGFEVIDVTKDGQDFVDVTLAVASEVNKDEQNLGIV 60

Query: 61  VDAYGVGPFMVATKVKGMIAAEVSDERSAYMTRAHNNARMITLGSEIVGPGVAKHIVEGF 120
           +DAYG GPFMVATK+KGM+AAEVSDERSAYMTR HNNARMIT+G+EIVG  +AK+I + F
Sbjct: 61  IDAYGAGPFMVATKIKGMVAAEVSDERSAYMTRGHNNARMITVGAEIVGDELAKNIAKAF 120

Query: 121 VDGTYDAGRHQIRVDMLNKMC 141
           V+G YD GRHQ+RVDMLNKMC
Sbjct: 121 VNGKYDGGRHQVRVDMLNKMC 141
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6911> which encodes the amino acid sequence <SEQ ID 6912>. Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
```

```
                 bacterial cytoplasm --- Certainty=0.3224(Affirmative) < succ>
                  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 101/140 (72%), Positives = 117/140 (83%)

Query: 1   MTIIIGADAHGVELKEVIRQHLTSLGKEIIDLTDTSKDFVDNTLAIVAKVNQKEDNLGIM 60
           M II+GADAHG  LKE+I+  L   G +IID+TD + DF+DNTLA+   VN+ E  LGIM
Sbjct: 1   MAIILGADAHGNALKELIKSFLQEEGYDIIDVTDINSDFIDNTLAVAKAVNEAEGRLGIM 60

Query: 61  VDAYGVGPFMVATKVKGMIAAEVSDERSAYMTRAHNNARMITLGSEIVGPGVAKHIVEGF 120
           VDAYG GPFMVATK+KGM+AAEVSDERSAYMTR HNNARMIT+G+EIVGP +AK+IV+GF
Sbjct: 61  VDAYGAGPFMVATKLKGMVAAEVSDERSAYMTRGHNNARMITIGAEIVGPELAKNIVKGF 120

Query: 121 VDGTYDAGRHQIRVDMLNKM 140
           V G YD GRHQIRVDMLNKM
Sbjct: 121 VTGPYDGGRHQIRVDMLNKM 140
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 2237**

A DNA sequence (GBSx2356) was identified in *S.agalactiae* <SEQ ID 6913> which encodes the amino acid sequence <SEQ ID 6914>. This protein is predicted to be galactose-6-phosphate isomerase lacb subunit (rpiB). Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2511(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10189> which encodes amino acid sequence <SEQ ID 10190> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAA25178 GB:M60447 galactose 6-P isomerase [Lactococcus lactis]
 Identities = 138/171 (80%), Positives = 157/171 (91%)

Query: 10  MKIAVGCDHIVTYDKIAVVDYLKTKGYEVIDCGTYDNIRTHYPIYGKKVGEAVASGKADL 69
           M+IA+GCDHIVT  K+AV ++LK+KGYEV+D GTYD++RTHYPIYGKKVGEAV SG+ADL
Sbjct: 1   MRIAIGCDHIVTDVKMAVSEFLKSKGYEVLDFGTYDHVRTHYPIYGKKVGEAVVSGQADL 60

Query: 70  GVCICGTGVGINNAVNKVPGIRSALVRDLTSAIYAKEELNANVIGFGGKITGGLLMTDII 129
           GVCICGTGVGINNAVNKVPG+RSALVRD+TSA+YAKEELNANVIGFGG ITGGLLM DII
Sbjct: 61  GVCICGTGVGINNAVNKVPGVRSALVRDMTSALYAKEELNANVIGFGGMITGGLLMNDII 120

Query: 130 EAFIRAKYKPTKENKVLIEKIAEVETHNAHQEENDFFTEFLDKWNRGEYHD 180
           EAFI A+YKPT+ENK LI KI  VETHNAHQ + +FFTEFL+KW+RGEYHD
Sbjct: 121 EAFIEAEYKPTEENKKLIAKIEHVETHNAHQADEEFFTEFLEKWDRGEYHD 171
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6915> which encodes the amino acid sequence <SEQ ID 6916>. Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3048(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 136/171 (79%), Positives = 160/171 (93%)

Query: 10  MKIAVGCDHIVTYDKIAVVDYLKTKGYEVIDCGTYDNIRTHYPIYGKKVGEAVASGKADL 69
            MKIA+GCDHIVT +K+AV D+LK+KGY+VIDCGTYD+ RTHYPI+GKKVGEAV +G+ADL
Sbjct: 2   MKIAIGCDHIVTNEKMAVSDFLKSKGYDVIDCGTYDHTRTHYPIFGKKVGEAVVNGQADL 61

Query: 70  GVCICGTGVGINNAVNKVPGIRSALVRDLTSAIYAKEELNANVIGFGGKITGGLLMTDII 129
            GVCICGTGVGINNAVNKVPGIRSALVRD+T+A+YAKEELNANVIGFGGKITG LLM DII
Sbjct: 62  GVCICGTGVGINNAVNKVPGIRSALVRDMTTALYAKEELNANVIGFGGKITGELLMCDII 121

Query: 130 EAFIRAKYKPTKENKVLIEKIAEVETHNAHQEENDFFTEFLDKWNRGEYHD 180
            +AFI+A+YK T+ENK LI KIA +E+H+A+QE+ DFFTEFL+KW+RGEYHD
Sbjct: 122 DAFIKAEYKETEENKKLIAKIAHLESHHANQEDPDFFTEFLEKWDRGEYHD 172
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 2238**

A DNA sequence (GBSx2357) was identified in *S.agalactiae* <SEQ ID 6917> which encodes the amino acid sequence <SEQ ID 6918>. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10187> which encodes amino acid sequence <SEQ ID 10188> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAA25179 GB:M60447 tagatose 6-P kinase [Lactococcus lactis]
 Identities = 192/310 (61%), Positives = 236/310 (75%)

Query: 11  MILTVTLNPSIDISYCLENFNMDTVNRVTDVSKTPGGKGLNVTRVLSQLGDNVVATGLLG 70
            MILTVTLNPS+DISY LE   +DTVNRV DVSKT GGKGLNVTRVL + GD V ATG LG
Sbjct: 1   MILTVTLNPSVDISYPLETLKIDTVNRVKDVSKTAGGKGLNVTRVLYESGDKVTATGFLG 60

Query: 71  GDFGDFIRSGLDALEIRHQFLSIGGETRHCIAVLHEGQQTEILEKGPHITKDEADAFLNH 130
            G  G+FI S L+   +   F  I G TR+CIA+LHEG QTEILE+GP I+ +EA+ FL+H
Sbjct: 61  GKIGEFIESELEQSPVSPAFYKISGNTRNCIAILHEGNQTEILEQGPTISHEEAEGFLDH 120

Query: 131 LKLIFDAATIITVSGSLPKGLPSDYYARLISLANHFNKKVVLDCSGEALRSVLKSSAKPT 190
               +   + ++T+SGSLP GLP+DYY +LI LA+     VVLDCSG  L +VLKSSAKPT
Sbjct: 121 YSNLIKQSEVVTISGSLPSGLPNDYYEKLIQLASDEGVAVVLDCSGAPLETVLKSSAKPT 180

Query: 191 VIKPNLEELTQLIGKPISYSLDELKSTLQQDLFRGIDWVIVSLGARGAFAKHGNHYYQVT 250
             IKPN EEL+QL+GK ++  ++ELK  L++ LF GI+W++VSLG  GAFAKHG+ +Y+V
Sbjct: 181 AIKPNNEELSQLLGKEVTKDIEELKDVLKESLFSGIEWIVVSLGRNGAFAKHGDVFYKVD 240

Query: 251 IPKIEVINPVGSGDATVAGIASALEHQLDDTNLLKRANVLGMLNAQETLTGHINLTYYQE 310
            IP I V+NPVGSGD+TVAGIASAL  +  D +LLK A  LGMLNAQET+TGH+N+T Y+
Sbjct: 241 IPDIPVVNPVGSGDSTVAGIASALNSKKSDADLLKHAMTLGMLNAQETMTGHVNMTNYET 300
```

```
Query: 311 LISQIQVKEV 320
           L SQI VKEV
Sbjct: 301 LNSQIGVKEV 310
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6919> which encodes the amino acid sequence <SEQ ID 6920>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1178(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 184/310 (59%), Positives = 232/310 (74%), Gaps = 1/310 (0%)

Query: 11  MILTVTLNPSIDISYCLENFNMDTVNRVTDVSKTPGGKGLNVTRVLSQLGDNVVATGLLG 70
            +ILTVTLNP+ID+SY L+    DTVNRV DV+KTPGGKGLNV+RVL++ G+ V ATG +G
Sbjct: 1   VILTVTLNPAIDVSYPLDELKCDTVNRVVDVTKTPGGKGLNVSRVLNEFGETVKATGCVG 60

Query: 71  GDFGDFIRSGLDALEIRHQFLSIGGETRHCIAVLHEGQQTEILEKGPHITKDEADAFLNH 130
            G+ GDFI + L    I  +F  I G+TR CIA+LHEG QTEILEKGP ++ DE D F +H
Sbjct: 61  GESGDFIINHLPD-SILSRFYKISGDTRTCIAILHEGNQTEILEKGPMLSVDEIDGFTHH 119

Query: 131 LKLIFDAATIITVSGSLPKGLPSDYYARLISLANHFNKKVVLDCSGEALRSVLKSSAKPT 190
             K + +   ++T+SGSLP G+P DYY +LI +AN   KK VLDCSG AL +VLK  +KPT
Sbjct: 120 FKYLLNDVDVVTLSGSLPAGMPDDYYQKLIKIANLNGKKTVLDCSGNALEAVLKGDSKPT 179

Query: 191 VIKPNLEELTQLIGKPISYSLDELKSTLQQDLFRGIDWVIVSLGARGAFAKHGNHYYQVT 250
            VIKPNLEEL+QL+GK ++   D LK  LQ +LF GI+W+IVSLGA G FAKH + +Y V
Sbjct: 180 VIKPNLEELSQLLGKEMTKDFDALKEVLQDELFDGIEWIIVSLGADGVFAKHKDTFYNVD 239

Query: 251 IPKIEVINPVGSGDATVAGIASALEHQLDDTNLLKRANVLGMLNAQETLTGHINLTYYQE 310
            IPKI++++ VGSGD+TVAGIAS L +  DD  LL +ANVLGMLNAQE  TGH+N+  Y +
Sbjct: 240 IPKIKIVSAVGSGDSTVAGIASGLANDEDDRALLTKANVLGMLNAQEKTTGHVNMANYDK 299

Query: 311 LISQIQVKEV 320
            L   I+VKEV
Sbjct: 300 LYQSIKVKEV 309
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 2239**

A DNA sequence (GBSx2358) was identified in *S.agalactiae* <SEQ ID 6921> which encodes the amino acid sequence <SEQ ID 6922>. This protein is predicted to be tagatose 1,6-diphosphate aldolase. Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0369(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAA25180 GB:M60447 tagatose 1,6-diP aldolase [Lactococcus
           lactis]
```

```
 Identities = 253/325 (77%), Positives = 295/325 (89%)

Query: 1   MGLTEQKQKHMEQLSDKNGIISALAFDQRGALKRLMAKYQSEEPTVSQIEALKVLVAEEL 60
           M LTEQK+K +E+LSDKNG ISALAFDQRGALKRLMA+YQ  EPTV+Q+E LKVLVA+EL
Sbjct: 1   MVLTEQKRKSLEKLSDKNGFISALAFDQRGALKRLMAQYQDTEPTVAQMEELKVLVADEL 60

Query: 61  TPYASSMLLDPEYGLPATKVLDDNAGLLLAYEKTGYDTSSTKRLPDCLDIWSAKRIKEEG 120
           T YASSMLLDPEYGLPATK LD  AGLLLA+EKTGYDTSSTKRLPDCLD+WSAKRIKE+G
Sbjct: 61  TKYASSMLLDPEYGLPATKALDKEAGLLLAFEKTGYDTSSTKRLPDCLDVWSAKRIKEQG 120

Query: 121 ADAVKFLLYYDVDSSDEVNEEKEAYIERIGSECVAEDIPFFLEILSYDEKITDSSGIEYA 180
           ADAVKFLLYYDVDSSDE+N++K+AYIER+GSECVAEDIPFFLEIL+YDE+I+D+  +EYA
Sbjct: 121 ADAVKFLLYYDVDSSDELNQQKQAYIERVGSECVAEDIPFFLEILAYDEEISDAGSVEYA 180

Query: 181 KIKPRKVIEAMKVFSNPRFNIDVLKVEVPVNMDYVEGFAQGETAYNKATAAAYFREQDQA 240
           K+KPRKVIEAMKVFS+PRFNIDVLKVEVPVN+ YVEGFA GE  Y+KA AA +F+ Q++A
Sbjct: 181 KVKPRKVIEAMKVFSDPRFNIDVLKVEVPVNVKYVEGFADGEVVYSKAEAADFFKAQEEA 240

Query: 241 TLLPYIFLSAGVPAQLFQETLVFAKEAGAKFNGVLCGRATWAGSVKEYVEKGEAGARQWL 300
           T LPYI+LSAGV A+LFQETL FA ++GAKFNGVLCGRATWAGSV+ Y+++GE  AR+WL
Sbjct: 241 TNLPYIYLSAGVSAKLFQETLQFAHDSGAKFNGVLCGRATWAGSVEPYIKEGEKAAREWL 300

Query: 301 RTIGFQNIDELNKILQKTATSWKER 325
           RT GF+NIDELNK+L KTA+ W ++
Sbjct: 301 RTTGFENIDELNKVLVKTASPWTDK 325
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6923> which encodes the amino acid sequence <SEQ ID 6924>. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0600(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 230/323 (71%), Positives = 276/323 (85%), Gaps = 1/323 (0%)

Query: 3   LTEQKQKHMEQLSDKNGIISALAFDQRGALKRLMAKYQSEEPTVSQIEALKVLVAEELTP 62
           LTE K+K ME+LS  +G+ISALAFDQRGALKR+MA++Q++EPTV QIE LK LV+EELTP
Sbjct: 5   LTENKRKSMEKLS-VDGVISALAFDQRGALKRMMAQHQTKEPTVEQIEELKSLVSEELTP 63

Query: 63  YASSMLLDPEYGLPATKVLDDNAGLLLAYEKTGYDTSSTKRLPDCLDIWSAKRIKEEGAD 122
           +ASS+LLDPEYGLPA++V   + AGLLLAYEKTGYD ++T RLPDCLD+WSAKRIKE GA+
Sbjct: 64  FASSILLDPEYGLPASRVRSEEAGLLLAYEKTGYDATTTSRLPDCLDVWSAKRIKEAGAE 123

Query: 123 AVKFLLYYDVDSSDEVNEEKEAYIERIGSECVAEDIPFFLEILSYDEKITDSSGIEYAKI 182
           AVKFLLYYD+D   +VNE+K+AYIERIGSEC AEDIPF+LEIL+YDEKI D++  E+AK+
Sbjct: 124 AVKFLLYYDIDGDQDVNEQKKAYIERIGSECRAEDIPFYLEILTYDEKIADNASPEFAKV 183

Query: 183 KPRKVIEAMKVFSNPRFNIDVLKVEVPVNMDYVEGFAQGETAYNKATAAAYFREQDQATL 242
           K  KV EAMKVFS  RF +DVLKVEVPVNM +VEGFA GE  + K  AA  FR+Q+ +T
Sbjct: 184 KAHKVNEAMKVFSKERFGVDVLKVEVPVNMKFVEGFADGEVLFTKEEAAQAFRDQEASTD 243

Query: 243 LPYIFLSAGVPAQLFQETLVFAKEAGAKFNGVLCGRATWAGSVKEYVEKGEAGARQWLRT 302
           LPYI+LSAGV A+LFQ+TLVFA E+GAKFNGVLCGRATWAGSVK Y+E+G   AR+WLRT
Sbjct: 244 LPYIYLSAGVSAKLFQDTLVFAAESGAKFNGVLCGRATWAGSVKVYIEEGPQAAREWLRT 303

Query: 303 IGFQNIDELNKILQKTATSWKER 325
            GF+NIDELNK+L KTA+ W E+
Sbjct: 304 EGFKNIDELNKVLDKTASPWTEK 326
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 2240**

A DNA sequence (GBSx2359) was identified in *S.agalactiae* <SEQ ID 6925> which encodes the amino acid sequence <SEQ ID 6926>. This protein is predicted to be lacx protein, chromosomal. Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0643(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10185> which encodes amino acid sequence <SEQ ID 10186> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAA25184 GB:M60447 ORF [Lactococcus lactis]
 Identities = 173/298 (58%), Positives = 219/298 (73%)

Query: 24  MAITIQNHELQVTLKALGATMTSITDSQGVEYLWQGDATYWGGQAPILFPICGSVRNDCV 83
            M I ++N  L V  K LG  +TSI D  G+EYLWQ D  YW GQAPILFPICGS+RND
Sbjct: 1   MTIELKNEYLTVQFKTLGGQLTSIKDKDGLEYLWQADPEYWNGQAPILFPICGSLRNDWA 60

Query: 84  IYRPAQAPHFTGIIPRHGFVRHKTFDYDYISDSSVRFTIKSSKEMLINYPYRFSLEITYT 143
            IYRP + P FTG+I RHGFVR + F  + ++++SV F+IK + EML NY Y+F L + YT
Sbjct: 61  IYRPQERPFFTGLIRRHGFVRKEEFTLEEVNENSVTFSIKPNAEMLDNYLYQFELRVVYT 120

Query: 144 LRNKSIAITYIVKNLESEKNMPYAIGAHPGFNCPLFEKEVFSDYYLEFEQFETCTIPESF 203
            L  KSI   + V NLE+EK MPY IGAHP FNCPL E E + DY LEF + E+C+IP+SF
Sbjct: 121 LNGKSIRTEFQVTNLETEKTMPYFIGAHPAFNCPLVEGEKYEDYSLEFSEVESCSIPKSF 180

Query: 204 PDTGLLDLQARHPFLENQKQLSLNHALFEKDAITLDQLRSKTVYLKSRNHAKGIQLDFDD 263
            P+TGLLDLQ R PFLENQK L L+++LF  DAITLD+L+S++V L+SR   KG+++DFDD
Sbjct: 181 PETGLLDLQDRTPFLENQKSLDLDYSLFSHDAITLDRLKSRSVTLRSRKSGKGLRVDFDD 240

Query: 264 FENLILWTSNNGGPFLALEPWSSLSTSIEESDILEDKQNIVRLNPKQSKQHSIRITIL 321
            F NLILW++ N  PF+ALEPWS LSTS+EE +ILEDK + ++  P  + + S  ITIL
Sbjct: 241 FPNLILWSTTNKSPFIALEPWSGLSTSLEEGNILEDKPQVTKVLPLDTSKKSYDITIL 298
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2241**

A DNA sequence (GBSx2361) was identified in *S.agalactiae* <SEQ ID 6927> which encodes the amino acid sequence <SEQ ID 6928>. This protein is predicted to be ABC transporter. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3272(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```

```
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10183> which encodes amino acid sequence <SEQ ID 10184> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA51350 GB:X72832 leucine rich protein [Streptococcus
           equisimilis]
 Identities = 101/278 (36%), Positives = 160/278 (57%), Gaps = 1/278 (0%)

Query: 10  MDFKELFPEVITKQEVKQSEDYIIVEQDGHVLHFPKSSLTKRELYLLQMTPSLEDASSVD 69
           M+ K+ FPE+         ++++ V++    +HFPKS L+++E  LL++     +
Sbjct: 1   MELKDYFPEMQVGPHPLGDKEWVSVKEGDQYVHFPKSCLSEKERLLLEVGLGQYEVLQ-P 59

Query: 70  SQNPWYRYLVEGRGRLPQSHSAVQFIFIEHQFTLSEELKDFLSPLVINVETIMTINQTQS 129
             +PW RYL++ +G  PQ     QFI++ HQ  L  +L + L  ++  +E I+ I+ TQ+
Sbjct: 60  LGSPWQRYLLDHQGNPPQLFETSQFIYLNHQQVLPADLVELLQQMIAGLEVILPISTTQT 119

Query: 130 VMILNQDNFFNATELLTDILPTIENDFNTRLRCYFGNSWTHLQAVDWKELYEEEYKLFTL 189
             +  Q         L  +LPT+E+DF   L  + GN+W  + A   +E +EEE +L T
Sbjct: 120 AFLCRQATSIKVLRSLEGLLPTLESDFGLALTMFVGNAWYQVAAGTLRECFEEECQLLTA 179

Query: 190 FLSHKAEQHYCRFPKMALWALANQSPMPSIKAKCLQHILDTSDTSAIIKALWQEQGNLAK 249
           +L  K+      F ++ LW++ +    P++  +  Q +   SD + ++ ALW E GNL +
Sbjct: 180 YLKQKSGGKLLTFAEVMLWSILSHQSFPALTRQFHQFLNPQSDMADVVHALWSEHGNLVQ 239

Query: 250 TAKALFIHRNSLQYKLDKFTQSSGLNLKILDDLAYAYL 287
           TA+ L+IHRNSLQYKLDKF Q SGL+LK LDDLA+AYL
Sbjct: 240 TAQRLYIHRNSLQYKLDKFAQQSGLHLKQLDDLAFAYL 277
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6929> which encodes the amino acid sequence <SEQ ID 6930>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.4332(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 106/287 (36%), Positives = 169/287 (57%), Gaps = 4/287 (1%)

Query: 3   KTVVED-AMDFKELFPEVITKQEVKQSEDYIIVEQDGHVLHFPKSSLTKRELYLLQM-TP 60
           KTV++  AM+ K+ FPE+         +D++ +++    +HFPKS L+++E  LL++
Sbjct: 7   KTVMKGMAMELKDYFPEMQVGPHPLGDKDWMSIKEGDQYVHFPKSCLSEKERLLLEVGLG 66

Query: 61  SLEDASSVDSQNPWYRYLVEGRGRLPQSHSAVQFIFIEHQFTLSEELKDFLSPLVINVET 120
             E    + S  PW RYL++ +G  PQ +   QFI++ HQ  L ++L + L  ++  +E
Sbjct: 67  QCEVLQPLGS--PWQRYLLDHQGNPPQLYETSQFIYLNHQQALPDDLVELLQQMIAGLEV 124

Query: 121 IMTINQTQSVMILNQDNFFNATELLTDILPTIENDFNTRLRCYFGNSWTHLQAVDWKELY 180
           I+ I+ TQ+  +  Q         L D+LPT+E+DF   L  + GN+W  + A   +E +
Sbjct: 125 ILPISATQTAFLCRQAISIKVLRWLEDLLPTLESDFGLALTMFVGNAWYQVAAGTLRECF 184

Query: 181 EEEYKLFTLFLSHKAEQHYCRFPKMALWALANQSPMPSIKAKCLQHILDTSDTSAIIKAL 240
           EEE +L T +L  ++ +    F  + LW+L +     ++  +  Q +   SD + ++ AL
Sbjct: 185 EEECQLLTAYLRQQSGRKLLTFSGLMLWSLLSHHTFLALTRQFHQFLSPQSDMADVVHAL 244

Query: 241 WQEQGNLAKTAKALFIHRNSLQYKLDKFTQSSGLNLKILDDLAYAYL 287
           W E GNL +TA+ L+IHRNSLQYKLDKF Q SGL+LK LDDLA+A+L
Sbjct: 245 WSEHGNLVQTAQRLYIHRNSLQYKLDKFAQQSGLHLKQLDDLAFAHL 291
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 2242**

A DNA sequence (GBSx2362) was identified in *S.agalactiae* <SEQ ID 6931> which encodes the amino acid sequence <SEQ ID 6932>. This protein is predicted to be multiple sugar-binding transport ATP-binding protein msmk (malK). Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4392(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAA26938 GB:M77351 ATP-binding protein [Streptococcus mutans]
 Identities = 320/377 (84%), Positives = 359/377 (94%)

Query: 1   MVELNLNHIYKKYPSASHYSVEDFDLDIKDKEFIVFVGPSGCGKSTTLRMIAGLEDISEG 60
            MVELNLNHIYKKYP++SHYSVEDFDLDIK+KEFIVFVGPSGCGKSTTLRM+AGLEDI++G
Sbjct: 1   MVELNLNHIYKKYPNSSHYSVEDFDLDIKNKEFIVFVGPSGCGKSTTLRMVAGLEDITKG 60

Query: 61  ELKIDGEVVNDKSPKDRDIAMVFQNYALYPHMTVYDNMAFGLKLRKFSKQEIDKRVREAA 120
            ELKIDGEVVNDK+PKDRDIAMVFQNYALYPHM+VYDNMAFGLKLR +SK+ IDKRV+EAA
Sbjct: 61  ELKIDGEVVNDKAPKDRDIAMVFQNYALYPHMSVYDNMAFGLKLRHYSKEAIDKRVKEAA 120

Query: 121 ANIGLTEFLERKPADLSGGQRQRVAMGRAIVRDAKVFLMDEPLSNLDAKLRVSMRAEIAK 180
              +GLTEFLERKPADLSGGQRQRVAMGRAIVRDAKVFLMDEPLSNLDAKLRVSMRAEIAK
Sbjct: 121 QILGLTEFLERKPADLSGGQRQRVAMGRAIVRDAKVFLMDEPLSNLDAKLRVSMRAEIAK 180

Query: 181 IHQRIGSTTIYVTHDQTEAMTLADRIVIMSATKNPDGDGTIGKIEQVGSPQELYNLPANK 240
            IH+RIG+TTIYVTHDQTEAMTLADRIVIMS+TKN DG GTIG++EQVG+PQELYN PANK
Sbjct: 181 IHRRIGATTIYVTHDQTEAMTLADRIVIMSSTKNEDGSGTIGRVEQVGTPQELYNRPANK 240

Query: 241 FVAGFIGSPSMNFFKVKVENGMIISEDGLRIAIPEGQEKLLESRGYKGKELIFGIRPEDI 300
            FVAGFIGSP+MNFF V +++G ++S+DGL IA+ EGQ K+LES+G+K K LIFGIRPEDI
Sbjct: 241 FVAGFIGSPAMNFFDVTIKDGHLVSKDGLTIAVTEGQLKMLESKGFKNKNLIFGIRPEDI 300

Query: 301 SSNLLVQDTYPNANVEAEVLVSELLGSETMLYVKLGQTEFASRVEARDFHNPGEKVNLTF 360
            SS+LLVQ+TYP+A V+AEV+VSELLGSETMLY+KLGQTEFA+RV+ARDFH PGEKV+LTF
Sbjct: 301 SSSLLVQETYPDATVDAEVVVSELLGSETMLYLKLGQTEFAARVDARDFHEPGEKVSLTF 360

Query: 361 NVAKGHFFDADTEQAIR 377
            NVAKGHFFDA+TE AIR
Sbjct: 361 NVAKGHFFDAETEAAIR 377
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6933> which encodes the amino acid sequence <SEQ ID 6934>. Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4642(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 332/377 (88%), Positives = 359/377 (95%)
```

```
Query: 1   MVELNLNHIYKKYPSASHYSVEDFDLDIKDKEFIVFVGPSGCGKSTTLRMIAGLEDISEG 60
           MVELNLNHIYKKYP+ +HY+VEDFDLDIKDKEFIVFVGPSGCGKSTTLRMIAGLEDISEG
Sbjct: 1   MVELNLNHIYKKYPNTTHYAVEDFDLDIKDKEFIVFVGPSGCGKSTTLRMIAGLEDISEG 60

Query: 61  ELKIDGEVVNDKSPKDRDIAMVFQNYALYPHMTVYDNMAFGLKLRKFSKQEIDKRVREAA 120
           ELKI GEVVNDKSPKDRDIAMVFQNYALYPHMTVYDNMAFGLKLRK+ K +ID+RV+EAA
Sbjct: 61  ELKIGGEVVNDKSPKDRDIAMVFQNYALYPHMTVYDNMAFGLKLRKYKKDDIDRRVKEAA 120

Query: 121 ANIGLTEFLERKPADLSGGQRQRVAMGRAIVRDAKVFLMDEPLSNLDAKLRVSMRAEIAK 180
             +GLTEFLERKPADLSGGQRQRVAMGRAIVRDAKVFLMDEPLSNLDAKLRVSMRAEIAK
Sbjct: 121 QILGLTEFLERKPADLSGGQRQRVAMGRAIVRDAKVFLMDEPLSNLDAKLRVSMRAEIAK 180

Query: 181 IHQRIGSTTIYVTHDQTEAMTLADRIVIMSATKNPDGDGTIGKIEQVGSPQELYNLPANK 240
           IH+RIGSTTIYVTHDQTEAMTLADRIVIMSATKNP G+GTIGKIEQVGSPQELYNLPANK
Sbjct: 181 IHRRIGSTTIYVTHDQTEAMTLADRIVIMSATKNPQGNGTIGKIEQVGSPQELYNLPANK 240

Query: 241 FVAGFIGSPSMNFFKVKVENGMIISEDGLRIAIPEGQEKLLESRGYKGKELIFGIRPEDI 300
           FVAGFIGSP+MNFF+V+V++G I+SEDGL IAIPEGQ K+LE+ GYKG+++ FGIRPEDI
Sbjct: 241 FVAGFIGSPAMNFFEVEVKDGRIVSEDGLDIAIPEGQAKMLEAAGYKGEKVTFGIRPEDI 300

Query: 301 SSNLLVQDTYPNANVEAEVLVSELLGSETMLYVKLGQTEFASRVEARDFHNPGEKVNLTF 360
           SS  +V DTYP+A V AEVLVSELLGSETMLYVKLGQTEFASRV+ARDFH+PGE+V+LTF
Sbjct: 301 SSRQIVHDTYPSATVTAEVLVSELLGSETMLYVKLGQTEFASRVDARDFHSPGEQVSLTF 360

Query: 361 NVAKGHFFDADTEQAIR 377
           NVAKGHFFD DTEQAIR
Sbjct: 361 NVAKGHFFDRDTEQAIR 377
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 2243**

A DNA sequence (GBSx2363) was identified in *S.agalactiae* <SEQ ID 6935> which encodes the amino acid sequence <SEQ ID 6936>. This protein is predicted to be glucan 1,6-alpha-glucosidase (dexB) (treC). Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2525(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA51348 GB:X72832 glucan 1,6-alpha-glucosidase [Streptococcus
           equisimilis]
 Identities = 413/535 (77%), Positives = 476/535 (88%), Gaps = 1/535 (0%)

Query: 1   MKKHWWHKATIYQIYPRSFMDSDGDGVGDIKGITSKLDYLEKLGITAIWLSPVYQSPMDD 60
           M+K WWHKATIYQIYPRSF D+ G+G+GD+KGITS+LDYL+KLGITAIWLSPVYQSPMDD
Sbjct: 1   MQKQWWHKATIYQIYPRSFKDTSGNGIGDLKGITSQLDYLQKLGITAIWLSPVYQSPMDD 60

Query: 61  NGYDISDYQAIADIFGDMNDMDQLLQEANQRGIKIIMDLVVNHTSDEHAWFVEARENPNS 120
           NGYDISDY+AIA++FG+M+DMD LL  AN+RGIKIIMDLVVNHTSDEHAWFVEARENPNS
Sbjct: 61  NGYDISDYEAIAEVFGNMDDMDDLLAAANERGIKIIMDLVVNHTSDEHAWFVEARENPNS 120

Query: 121 PERDFYIWRDEPNDLTSIFSGSAWEYDKVSGQYYLHLFSKRQPDLNWENEALRHKIYDMM 180
           PERD+YIWRDEPN+L SIFSGSAWE D+ SGQYYLHLFSK+QPDLNWEN  +R KIYDMM
Sbjct: 121 PERDYYIWRDEPNNLMSIFSGSAWELDEASGQYYLHLFSKKQPDLNWENAHVRQKIYDMM 180

Query: 181 NFWIDKGIGGFRMDVIDLIGKIPDKGITGNGPKLHDYLKEMNRASFGKHDLLTVGETWGA 240
```

```
           NFWI KGIGGFRMDVIDLIGKIPD  ITGNGP+LHDYLKEMN+A+FG HD++TVGETWGA
Sbjct: 181 NFWIAKGIGGFRMDVIDLIGKIPDSEITGNGPRLHDYLKEMNQATFGNHDVMTVGETWGA 240

Query: 241 TPDIAKQYSNPDNEELSMVFQFEHVGLQHKPDAPKWDYSDGLDVPALKDIFTKWQTQLEL 300
           TP+IA+QYS P+N+ELSMVFQFEHVGLQHKP+APKWDY++ LDVPALK IF+KWQT+L+L
Sbjct: 241 TPEIARQYSRPENKELSMVFQFEHVGLQHKPNAPKWDYAEELDVPALKTIFSKWQTELKL 300

Query: 301 GQGWNSLFWNNHDLPRVLSIWGNDSDNRKQSAKALAILLHLMRGTPYIYQGEEIGMTNYP 360
           G+GWNSLFWNNHDLPRVLSIWGNDS  R++SAKALAILLHLMRGTPYIYQGEEIGMTNYP
Sbjct: 301 GEGWNSLFWNNHDLPRVLSIWGNDSIYREKSAKALAILLHLMRGTPYIYQGEEIGMTNYP 360

Query: 361 FECLADVDDIESLNYAKEAMDNGVSEATILDSIRKVGRDNARTPMQWSQEHQAGFTKG-T 419
           F+ L +VDDIESLNYAKEAM+NGV  A ++ SIRKVGRDNARTPMQWS++  AGF++
Sbjct: 361 FKDLTEVDDIESLNYAKEAMENGVPAARVMSSIRKVGRDNARTPMQWSKDTHAGFSEAQE 420

Query: 420 PWLAVNPNYQEINVEAALNDTESIFYTYQKLVALRKEHDWLVDADFKLLETADKVFAYVR 479
            WL VNPNYQEINV  AL + +SIFYTYQ+L+ALRK+ DWLV+AD+ LL TADKVFAY R
Sbjct: 421 TWLPVNPNYQEINVADALANQDSIFYTYQQLIALRKDQDWLVEADYHLLPTADKVFAYQR 480

Query: 480 QTDKERYLIVANLSDQNQSFEFPEAVKETIISNTEVQEVLSSNTLKPWDAFCIEL 534
           Q  +E Y+IV N+SDQ Q F    A  E +I+NT+V +VL +  L+PWDAFC++L
Sbjct: 481 QFGEETYVIVVNVSDQEQVFAKDLAGAEVVITNTDVDKVLETKHLQPWDAFCVKL 535
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6937> which encodes the amino acid sequence <SEQ ID 6938>. Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2793(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 418/535 (78%), Positives = 474/535 (88%), Gaps = 1/535 (0%)

Query: 1   MKKHWWHKATIYQIYPRSFMDSDGDGVGDIKGITSKLDYLEKLGITAIWLSPVYQSPMDD 60
           M  HWWHKATIYQIYPRSF D+ G+G+GD+KGITS+LDYL+KLGITAIWLSPVYQSPMDD
Sbjct: 1   MNNHWWHKATIYQIYPRSFKDTSGNGIGDLKGITSQLDYLQKLGITAIWLSPVYQSPMDD 60

Query: 61  NGYDISDYQAIADIFGDMNDMDQLLQEANQRGIKIIMDLVVNHTSDEHAWFVEARENPNS 120
           NGYDISDY+AIAD+FGDM DMD+LL  AN+RGIKIIMDLVVNHTSDEHAWFVEARENPNS
Sbjct: 61  NGYDISDYEAIADVFGDMADMDELLAAANERGIKIIMDLVVNHTSDEHAWFVEARENPNS 120

Query: 121 PERDFYIWRDEPNDLTSIFSGSAWEYDKVSGQYYLHLFSKRQPDLNWENEALRHKIYDMM 180
           PERD+YIWRDEPN+L SIFSGSAWE D+ SGQYYLHLFSK+QPDLNWEN  LR KIYDMM
Sbjct: 121 PERDYYIWRDEPNNLMSIFSGSAWELDEASGQYYLHLFSKKQPDLNWENAQLRQKIYDMM 180

Query: 181 NFWIDKGIGGFRMDVIDLIGKIPDKGITGNGPKLHDYLKEMNRASFGKHDLLTVGETWGA 240
           NFWI KGIGGFRMDVIDLIGK+PD  ITGNGP+LHDYLKEMN+A+FG HD++TVGETWGA
Sbjct: 181 NFWIAKGIGGFRMDVIDLIGKVPDLEITGNGPRLHDYLKEMNQATFGNHDVMTVGETWGA 240

Query: 241 TPDIAKQYSNPDNEELSMVFQFEHVGLQHKPDAPKWDYSDGLDVPALKDIFTKWQTQLEL 300
           TP+IA+QYS P+N+ELSMVFQFEHVGLQHKPDAPKWDY+  LDVPALK IF+KWQT+L+L
Sbjct: 241 TPEIARQYSRPENKELSMVFQFEHVGLQHKPDAPKWDYAKELDVPALKAIFSKWQTELKL 300

Query: 301 GQGWNSLFWNNHDLPRVLSIWGNDSDNRKQSAKALAILLHLMRGTPYIYQGEEIGMTNYP 360
           G+GWNSLFWNNHDLPRVLSIWGNDS  R++SAKALAILLHLMRGTPYIYQGEEIGMTNYP
Sbjct: 301 GEGWNSLFWNNHDLPRVLSIWGNDSTYREKSAKALAILLHLMRGTPYIYQGEEIGMTNYP 360

Query: 361 FECLADVDDIESLNYAKEAMDNGVSEATILDSIRKVGRDNARTPMQWSQEHQAGFTKG-T 419
           F+ L +V+DIESLNYAKEAM NGVS A ++DSIRKVGRDNARTPMQWS++  AGF++
Sbjct: 361 FKDLTEVNDIESLNYAKEAMGNGVSAARVMDSIRKVGRDNARTPMQWSKDTHAGFSEAKE 420

Query: 420 PWLAVNPNYQEINVEAALNDTESIFYTYQKLVALRKEHDWLVDADFKLLETADKVFAYVR 479
```

```
            WL VNPNYQ+INV  AL D +SIFYTYQKL+ALRKE DWLV+AD+ LL TADKVFAY R
Sbjct: 421 TWLPVNPNYQDINVADALADPDSIFYTYQKLIALRKEQDWLVEADYHLLPTADKVFAYQR 480

Query: 480 QTDKERYLIVANLSDQNQSFEFPEAVKETIISNTEVQEVLSSNTLKPWDAFCIEL 534
           Q  +E Y+IV N+SD+ Q F    A  + II+NT+V  VL +  L+PWDAFC++L
Sbjct: 481 QLGEETYVIVVNVSDEEQVFATDLAGAQVIIANTDVDTVLETKHLQPWDAFCLKL 535
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## Example 2244

A DNA sequence (GBSx2364) was identified in *S.agalactiae* <SEQ ID 6939> which encodes the amino acid sequence <SEQ ID 6940>. Analysis of this protein sequence reveals the following:

```
Possible site: 44
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAB49738 GB:U21942 UDP-galactose 4-epimerase [Streptococcus mutans]
 Identities = 267/331 (80%), Positives = 306/331 (91%)

Query: 1   MAVLILGGAGYIGSHMVDQLITQGKEKVIVVDNLVTGHRQAVHSDAIFYEGDLSDKTFMR 60
           MA+L+LGGAGYIGSHMVD+LI +G+E+V+VVD+LVTGHR AVH  A FY+GDL+D+ FM
Sbjct: 1   MAILVLGGAGYIGSHMVDRLIEKGEEEVVVVDSLVTGHRAAVHPAAKFYQGDLADREFMS 60

Query: 61  QVFRENPDVDAVIHFAAFSLVAESMENPLKYFDNNTAGMIKLLEVMNECDIKNIVFSSTA 120
            VFRENPDVDAVIHFAA+SLVAESM+ PLKYFDNNTAGMIKLLEVM+E  +K IVFSSTA
Sbjct: 61  MVFRENPDVDAVIHFAAYSLVAESMKKPLKYFDNNTAGMIKLLEVMSEFGVKYIVFSSTA 120

Query: 121 ATYGIPEQVPILETAPQNPINPYGESKLMMETIMKWADQAYGIKFVALRYFNVAGDKPDG 180
           ATYGIP ++PI ET PQ PINPYGESKLMMETIMKW+D+AYGIKFV +RYFNVAG KPDG
Sbjct: 121 ATYGIPNEIPIKETTPQRPINPYGESKLMMETIMKWSDRAYGIKFVPVRYFNVAGAKPDG 180

Query: 181 SIGEDHKPETHLLPIILQVAQGVRDKIMIFGDDYNTPDGTNVRDYVHPFDLADAHILAVD 240
           SIGEDH PETHLLPIILQVAQGVR+KIMIFGDDYNTPDGTNVRDYVHPFDLAD H+LA++
Sbjct: 181 SIGEDHSPETHLLPIILQVAQGVREKIMIFGDDYNTPDGTNVRDYVHPFDLADRHLLALN 240

Query: 241 YLRQGNESNVFNLGSSTGFSNLQMLEAARRITGKEIPAQKAARRPGDPDTLIASSEKARQ 300
           YLRQGN S  FNLGSSTGFSNLQ+LEAAR++TG++IPA+KAARR GDPDTLIASSEKAR+
Sbjct: 241 YLRQGNPSTAFNLGSSTGFSNLQILEAARKVTGQKIPAEKAARRSGDPDTLIASSEKARE 300

Query: 301 ILGWEPKFDNIDKIISSAWAWHSSHPNGYED 331
           ++GW+P+FD+I+KII+SAWAWHSSHP GY+D
Sbjct: 301 VVGWKPQFDDIEKIIASAWAWHSSHPKGYDD 331
```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

## Example 2245

A DNA sequence (GBSx2366) was identified in *S.agalactiae* <SEQ ID 6941> which encodes the amino acid sequence <SEQ ID 6942>. This protein is predicted to be two-component response regulator. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3945(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB06470 GB:AP001516 two-component response regulator [Bacillus halodurans]
 Identities = 71/223 (31%), Positives = 139/223 (61%), Gaps = 7/223 (3%)

Query: 3   VLIIEDDPMVEFIHRNYLEKLNYFQNIYSTASQTQAIAYLNDIKIQLVLLDIHIKEGNGL 62
           VL+IEDDPMV+ ++R ++EKL+ F  + +TA+  + +    +++  L+LLDI + + +GL
Sbjct: 9   VLLIEDDPMVQEVNRMFVEKLSGFTIVGTTATGEEGMVKTRELQPDLILLDIFMPKQDGL 68

Query: 63  ELLKLLRNQHQNTEVIVISAANEAHTVKEAFHLGIVDYLIKPFTFERFESSIEKFLNHYH 122
             +K +R Q+ + ++I ++AAN+  T+K     G++DYL+KPFTFER ++++ ++   +
Sbjct: 69  SFIKQIREQYIDVDIIAVTAANDTKTIKTLLRYGVMDYLVKPFTFERLKAALTQYEEMFR 128

Query: 123 TFEAD-KIYQDNIDHFQKIDSGWLEGEVKLDE--KGLSEITYQHILDAIQELEQPFTIQE 179
             + + ++ QD++D   K      + +  +D+  KGL   T Q +++ ++EL++P + +E
Sbjct: 129 KMQKEAELSQDSLDEMIK----QKQAQANMDDLPKGLHAHTLQQVIERLEELDEPKSAEE 184

Query: 180 LAKCSQFSHVSVRKYIAYMEEKGLLTSQQIYTKVGRPYKVYKL 222
           + +    + V+VR+Y+ Y+E  G +     Y  +GRP + YKL
Sbjct: 185 IGRDVGLARVTVRRYLNYLESVGQVEMDLTYGSIGRPIQTYKL 227
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6943> which encodes the amino acid sequence <SEQ ID 6944>. Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4053(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 123/220 (55%), Positives = 156/220 (70%)

Query: 1   MDVLIIEDDPMVEFIHRNYLEKLNYFQNIYSTASQTQAIAYLNDIKIQLVLLDIHIKEGN 60
           M+VLIIEDDPMV+FIHRNYLEKLN F  I S+ S     + L D  I L+LLDIHI +GN
Sbjct: 1   MNVLIIEDDPMVDFIHRNYLEKLNLFDRIISSDSMKAVQSILTDYAIDLILLDIHITDGN 60

Query: 61  GLELLKLLRNQHQNTEVIVISAANEAHTVKEAFHLGIVDYLIKPFTFERFESSIEKFLNH 120
           G++ L+  R QH   EVI+ISAAN+ + +++ FHLGI+DYLIKPFTFERF+ SI++F+ H
Sbjct: 61  GIQFLEKWRTQHIPCEVIIISAANDGNIIRDGFHLGIIDYLIKPFTFERFQESIQQFVTH 120

Query: 121 YHTFEADKIYQDNIDHFQKIDSGWLEGEVKLDEKGLSEITYQHILDAIQELEQPFTIQEL 180
                  ++ Q  ID  + + S       +L EKGLSE T+Q I++ I+  +QPFTIQEL
Sbjct: 121 REHLANQQLEQAQIDQLKCLTSKKDTKNKQLLEKGLSESTFQWIMENIKVFDQPFTIQEL 180

Query: 181 AKCSQFSHVSVRKYIAYMEEKGLLTSQQIYTKVGRPYKVY 220
           A     SHVSVRKYIAY+EE   L SQQI+TKVGRPY+VY
Sbjct: 181 ASACHLSHVSVRKYIAYLEENKQLNSQQIFTKVGRPYRVY 220
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 2246

A DNA sequence (GBSx2367) was identified in *S.agalactiae* <SEQ ID 6945> which encodes the amino acid sequence <SEQ ID 6946>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -8.76   Transmembrane   12 -  28 (   6 -  34)
   INTEGRAL    Likelihood = -7.43   Transmembrane  178 - 194 ( 173 - 197)

----- Final Results -----
              bacterial membrane --- Certainty=0.4503(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9003> which encodes amino acid sequence <SEQ ID 9004> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 3
SRCFLG: 0
McG: Length  of  UR:    27
     Peak Value of UR:    2.99
     Net Charge of CR: 3
McG: Discrim Score:      12.92
GvH: Signal Score (-7.5): -2.57
     Possible site: 19
>>> Seems to have an uncleavable N-term signal seq
Amino Acid Composition: calculated from 1
ALOM program   count: 2 value:  -8.76 threshold:  0.0
   INTEGRAL    Likelihood = -8.76   Transmembrane   10 -  26 (   4 -  32)
   INTEGRAL    Likelihood = -7.43   Transmembrane  176 - 192 ( 171 - 195)
   PERIPHERAL  Likelihood =  3.18     149
 modified ALOM score:   2.25
icm1 HYPID: 7  CFP: 0.450

*** Reasoning Step: 3

----- Final Results -----
              bacterial membrane --- Certainty=0.4503(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB15141 GB:Z99120 similar to two-component sensor histidine
           kinase [YufM] [Bacillus subtilis]
 Identities = 132/461 (28%), Positives = 245/461 (52%), Gaps = 7/461 (1%)

Query: 3   MKKKLSLWAFLSLILVTMTICIFSIFYYVTIHQSYRMVRVQEEKILKNTGYALSRNPQVI 62
           MKK L L   L++ +   + +   I ++    Q+ + +R QE+     T   ++  P
Sbjct: 1   MKKTLKLQTRLTIFVCIVVLIALLITFFTVGAQTTKRIRDQEKATALQTAEMVAEAPMTA 60

Query: 63  QTLKDNHYDQSLQKQMLFLSKKSNLDYIVLINLKGIRFTHPDSTKIGKPFQGGDEQAVFK 122
             L+     + LQ     + K +  +++V++++ GIR THPD +KIGK F+GGDE  V K
Sbjct: 61  AALESGKKQKELQSYTKRVQKITGTEFVVVMDMNGIRKTHPDPSKIGKKFRGGDESEVLK 120

Query: 123 GKAIMSTAEGSLGKSLRYLIPVY-DHQKQVGAIAVGLKLTTLGDLSQSSIKEFSKPLLIS 181
           G   +STA G+LGKS R  +PVY ++ KQVGA+AVG+ +  + ++   S++     + +S
Sbjct: 121 GHVHISTASGTLGKSQRAFVPVYAENGKQVGAVAVGITVNEIDEVISHSLRPLYFIICVS 180

Query: 182 ILISLVVTSIISYGLKKQLHNLHPSDIFQHLEERNATLDQIQAAVFVIDQRHIIKRNNPA 241
           I + ++   I++  +K  ++ L P +I   LEER+A L+  +  +  +D+   IK  N
Sbjct: 181 IFVGVIGAVIVARTVKNIMYGLEPYEIATLLEERSAMLESTKEGILAVDEHGKIKLANAE 240

Query: 242 ASLLFKKEGQRDLFSGKLLESLIP--QLKQDHFSKK--TEQVLHFQGQDYLLSISPITVK 297
           A  LF K G       + ++ ++P  +LK+   +KK   ++ +   G + + +  PI +K
Sbjct: 241 AKRLFVKMGINTNPIDQDVDDILPKSRLKKVIETKKPLQDRDVRINGLELVFNEVPIQLK 300

Query: 298 TQNRGYVVFLRNVTETLFTLDQLAHTTAYASALQAQTHQFMNQLHVIYGLADIEYYDELK 357
            Q  G +   R+ TE     +QL+    YA+AL+AQ+H+FMN+LHVI GL  ++ YD+L
Sbjct: 301 GQTVGAIATFRDKTEVKHLAEQLSGVKMYANALRAQSHEFMNKLHVILGLVQLKEYDDLG 360
```

```
Query: 358 IYLKELLEPQNEFLARLSMLVREPRLASFIIGEREKFAEKHINLSTEILVEIPTKSTVED 417
            Y+K++   Q     + +  V+   LA F++G++    E+  NL  E    IP  +
Sbjct: 361 DYIKDIAIQQKSETSEIINDVKSSVLAGFLLGKQSFIREQGANLDIECNGVIPNAADPSV 420

Query: 418 VNNYL-LLHRYINTKILTLLN-STTLVSLRLNYQNNLIETD 456
           ++   + ++    IN  +  + +     +++ + + N++++ +
Sbjct: 421 IHELITIIGNLINNGLDAVADMPKKQITMSMRFHNSILDIE 461
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6947> which encodes the amino acid sequence <SEQ ID 6948>. Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-10.03   Transmembrane  174 - 190 ( 170 - 195)

----- Final Results -----
               bacterial membrane --- Certainty=0.5012(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 236/488 (48%), Positives = 337/488 (68%), Gaps = 3/488 (0%)

Query: 3   MKKKLSLWAFLSLILVTMTICIFSIFYYVTIHQSYRMVRVQEEKILKNTGYALSRNPQVI 62
           MKK L LWA LSLILV+M +   S+FY + +H +++ ++ QE  +L +TG  L+ +  +
Sbjct: 1   MKKPLRLWASLSLILVSMIVVTTSLFYGIMLHDTHQSIKNQETHLLTSTGKMLASHQAIK 60

Query: 63  QTLKDNHYDQSLQKQMLFLSKKSNLDYIVLINLKGIRFTHPDSTKIGKPFQGGDEQAVFK 122
           + L +N  +         ++   NLDY+V++N+KGIR THP+   IGKPFQGGDE+AV
Sbjct: 61  ELLLNNQPNAKTTAYTNSIASIYNLDYVVVMNMKGIRLTHPNPKNIGKPFQGGDEEAVLA 120

Query: 123 GKAIMSTAEGSLGKSLRYLIPVYDHQKQVGAIAVGLKLTTLGDLSQSSIKEFSKPLLISI 182
           GK ++STA+G+LGKSLRYL+PV+D  KQ+GAIAVG+KLTTL D++ +S + ++  LL+ +
Sbjct: 121 GKKVISTAKGTLGKSLRYLVPVFDGDKQIGAIAVGIKLTTLNDVALTSKRNYTLSLLLCL 180

Query: 183 LISLVVTSIISYGLKKQLHNLHPSDIFQHLEERNATLDQIQAAVFVIDQRHIIKRNNPAA 242
           LISL+VTS IS+ LK+QLH L PS+I+Q  EERNA LDQI+AAVFV+D+  I++  N A
Sbjct: 181 LISLLVTSFISFRLKRQLHQLEPSEIYQLFEERNAMLDQIEAAVFVVDKAGILQLCNQAG 240

Query: 243 SLLFKKEGQRDLFSGKLLESLIPQLKQDHFSKKTEQVLHFQGQDYLLSISPITVKTQNRG 302
             L  ++ Q    +G     L P    +     +  EQ+  +  +DYLL+ISPI VK  +RG
Sbjct: 241 QKLIARKCQLGKPTGNSFNYLFPDFPKLSLQEGHEQLFRYGEEDYLLAISPICVKNDHRG 300

Query: 303 YVVFLRNVTETLFTLDQLAHTTAYASALQAQTHQFMNQLHVIYGLADIEYYDELKIYLKE 362
           +++F+R   + + TLDQLA+TTAYASALQAQTH+FMNQLHVIYGL DI YYD+LKIYL
Sbjct: 301 HIIFMREAVKAIDTLDQLAYTTAYASALQAQTHKFMNQLHVIYGLVDIAYYDQLKIYLDS 360

Query: 363 LLEPQNEFLARLSMLVREPRLASFIIGEREKFAEKHINLSTEILVEIPTKSTVEDVNNYL 422
           +LEP+NE L  LS+LV+EP LASF+IGE+EK+ E +++L  ++L EIP  +T   +NN L
Sbjct: 361 ILEPENEILTSLSVLVKEPLLASFLIGEQEKYQELNVHLKIDVLSEIPHSATKNQLNNGL 420

Query: 423 LLHRYINTKILTLLNSTTLVSLRLNYQNNLIETDYQWENEKWL-LNDYHQYFNDAYFQQL 481
           +++R+I+T +LT L   +LV    + QN+LI    +    + W+ L      F+  YFQQL
Sbjct: 421 MIYRFIHTNLLTTLRPKSLVLSIQHDQNHLI--SHYTLTDNWIDLERVQPIFDLPYFQQL 478

Query: 482 LVDSRATY 489
           L D+ + +
Sbjct: 479 LTDTNSQF 486
```

SEQ ID 9004 (GBS130d) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 123 (lane 8-10; MW 63kDa) and in Figure 184 (lane 4; MW 63kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 123 (lane 11; MW 38kDa) and in Figure 181 (lane 7; MW 38kDa).

GBS130d-GST was purified as shown in Figure 237, lane 11. GBS130d-His was purified as shown in Figure 233, lane 9-10.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## Example 2247

A DNA sequence (GBSx2368) was identified in *S.agalactiae* <SEQ ID 6949> which encodes the amino acid sequence <SEQ ID 6950>. Analysis of this protein sequence reveals the following:

```
Possible site: 51
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-11.52   Transmembrane  364 - 380 ( 353 - 386)
   INTEGRAL    Likelihood = -9.66   Transmembrane   33 -  49 (  26 -  57)
   INTEGRAL    Likelihood = -7.80   Transmembrane   87 - 103 (  82 - 105)
   INTEGRAL    Likelihood = -6.85   Transmembrane  153 - 169 ( 144 - 174)
   INTEGRAL    Likelihood = -4.41   Transmembrane  301 - 317 ( 300 - 318)
   INTEGRAL    Likelihood = -2.81   Transmembrane  216 - 232 ( 212 - 235)
   INTEGRAL    Likelihood = -2.39   Transmembrane  120 - 136 ( 120 - 136)
   INTEGRAL    Likelihood = -1.65   Transmembrane   57 -  73 (  56 -  73)
   INTEGRAL    Likelihood = -1.17   Transmembrane  428 - 444 ( 428 - 444)
   INTEGRAL    Likelihood = -0.32   Transmembrane  276 - 292 ( 276 - 292)

----- Final Results -----
               bacterial membrane --- Certainty=0.5607(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAB18291 GB:U35658 L-malate permease [Streptococcus bovis]
 Identities = 329/428 (76%), Positives = 375/428 (86%)

Query: 18  DLKAKLFHIKIGSVPLPVYVCLALLILLAGFLQKLPVNMLGGFAVILTMGWFLGTIGASI 77
            D + KL   +IGSV LPVY+  A +IL+   L++LPVNMLGGFAVILTMGW LGTIG +I
Sbjct: 14  DWRNKLTKTRIGSVTLPVYLVTASIILVTALLEQLPVNMLGGFAVILTMGWLLGTIGGNI 73

Query: 78  PGFKNFGGPAILSLLVPSILVFFNLINKNVLESTNMLMKQANFLYFYIACLVSGSILGMN 137
            P  K+FGGPAILSLLVPSI+VFFNL+N+NVL+ST++LMKQANFLYFYIACLV GSILGMN
Sbjct: 74  PILKHFGGPAILSLLVPSIMVFFNLLNQNVLDSTDILMKQANFLYFYIACLVCGSILGMN 133

Query: 138 RKMLIQGLLRMIFPMLLGMVCAMMVGTFVGVILGLEWRHTLFYIVTPVLAGGIGEGILPL 197
            RK+L+QGL+RMI PM LGM+ AM VGT VG +LGL W+H+LFYIVTPVLAGGIGEGILPL
Sbjct: 134 RKILVQGLMRMIVPMALGMILAMGVGTLVGTLLGLGWKHSLFYIVTPVLAGGIGEGILPL 193

Query: 198 SLGYSSITGVASEQLVAQLIPATIIGNFFAILCTALLNRLGEKKPHLSGQGQLVRLNKGE 257
            SLGYS+ITG+ SEQLV QLIPATIIGNFFAI+C+ LL+RLGEK+P LSGQGQL+++   +
Sbjct: 194 SLGYSAITGLPSEQLVGQLIPATIIGNFFAIMCSGLLSRLGEKRPELSGQGQLIKITNSD 253

Query: 258 DMSDIIADHSGPIDVKKMGGGVLTACSLFIFGHLLQQLTGFPGPVLMIVAAAILKYINVI 317
            D+SD + +   PIDVK MG GVL AC+LFI G LLQ LTGFPGPVLMIV AA LKY+NV+
Sbjct: 254 DLSDALEEDKAPIDVKLMGAGVLIACTLFITGGLLQHLTGFPGPVLMIVVAAFLKYLNVV 313

Query: 318 PRETQNGAKQLYKFISGNFTFPLMAGLGLLYIPLKDVVATLSIQYFIVVISVVFTVISVG 377
            P+ETQ G+KQLYKFISGNFTFPLM GLG+LYIPLKDVV  LS QYF+VVISVVFTVI+ G
Sbjct: 314 PKETQRGSKQLYKFISGNFTFPLMVGLGMLYIPLKDVVGMLSWQYFVVVISVVFTVIATG 373

Query: 378 FFVSRFLNMNPVEAGIISACQSGMGGTGDVAILSTADRMNLMPFAQVATRLGGAITVITM 437
            FFVSRF+NMNPVEA I+SACQSGMGGTGDVAILSTA+RM LMPFAQVATRLGGAITVITM
Sbjct: 374 FFVSRFMNMNPVEAAIVSACQSGMGGTGDVAILSTANRMTLMPFAQVATRLGGAITVITM 433

Query: 438 TAILRMLF 445
            TAI RMLF
Sbjct: 434 TAIFRMLF 441
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6951> which encodes the amino acid sequence <SEQ ID 6952>. Analysis of this protein sequence reveals the following:

```
     Possible site: 48
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-11.89   Transmembrane  361 - 377 ( 350 - 383)
   INTEGRAL    Likelihood = -7.43   Transmembrane   84 - 100 (  79 - 102)
   INTEGRAL    Likelihood = -6.16   Transmembrane  150 - 166 ( 137 - 171)
   INTEGRAL    Likelihood = -4.88   Transmembrane   30 -  46 (  24 -  48)
   INTEGRAL    Likelihood = -4.35   Transmembrane  299 - 315 ( 297 - 316)
   INTEGRAL    Likelihood = -4.14   Transmembrane  117 - 133 ( 115 - 134)
   INTEGRAL    Likelihood = -3.19   Transmembrane   54 -  70 (  51 -  75)
   INTEGRAL    Likelihood = -2.92   Transmembrane  425 - 441 ( 425 - 442)
   INTEGRAL    Likelihood = -2.81   Transmembrane  213 - 229 ( 209 - 232)
   INTEGRAL    Likelihood = -2.44   Transmembrane  273 - 289 ( 271 - 290)

----- Final Results -----
               bacterial membrane --- Certainty=0.5755(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAB18291 GB:U35658 L-malate permease [Streptococcus bovis]
 Identities = 344/443 (77%), Positives = 394/443 (88%), Gaps = 6/443 (1%)

Query: 4   ISKKMPQKDLSEHSKAWQNR----RIGSVPLPVYLVLATLILVTGWLQQLPVNMLGGFAV 59
           + KK+P    +E    W+N+    RIGSV LPVYLV A++ILVT  L+QLPVNMLGGFAV
Sbjct: 1   MEKKLPATAANETD--WRNKLTKTRIGSVTLPVYLVTASIILVTALLEQLPVNMLGGFAV 58

Query: 60  ILTLGWLLGTIGATIPGLKHFGGPAILSLLVPSILVFFNLLNPNVLEATNVLMKQANFLY 119
           ILT+GWLLGTIG  IP LKHFGGPAILSLLVPSI+VFFNLLN NVL++T++LMKQANFLY
Sbjct: 59  ILTMGWLLGTIGGNIPILKHFGGPAILSLLVPSIMVFFNLLNQNVLDSTDILMKQANFLY 118

Query: 120 FYIACLVCGSILGMNRKILIQGLFRMIIPMLLGMVCAMGVGTLVGVILGLDWQHTLFYVV 179
           FYIACLVCGSILGMNRKIL+QGL RMI+PM LGM+ AMGVGTLVG +LGL W+H+LFY+V
Sbjct: 119 FYIACLVCGSILGMNRKILVQGLMRMIVPMALGMILAMGVGTLVGTLLGLGWKHSLFYIV 178

Query: 180 TPVLAGGIGEGILPLSLGYSAITGVGSEQLVAQLIPATIIGNFFAILCTALLNRFGEKHP 239
           TPVLAGGIGEGILPLSLGYSAITG+ SEQLV QLIPATIIGNFFAI+C+ LL+R GEK P
Sbjct: 179 TPVLAGGIGEGILPLSLGYSAITGLPSEQLVGQLIPATIIGNFFAIMCSGLLSRLGEKRP 238

Query: 240 SYSGQGQLVKIGHSEDMSDALKDNSGALDVKLMGAGVLTACSLFIAGGLLQHLTDFPGPV 299
             SGQGQL+KI +S+D+SDAL+++   +DVKLMGAGVL AC+LFI GGLLQHLT FPGPV
Sbjct: 239 ELSGQGQLIKITNSDDLSDALEEDKAPIDVKLMGAGVLIACTLFITGGLLQHLTGFPGPV 298

Query: 300 LMIILAAFLKYLNVIPQETQNGAKQLYKFISSNFTFPLMAGLGLLYIPLKEVVATLSWQY 359
           LMI++AAFLKYLNV+P+ETQ G+KQLYKFIS NFTFPLM GLG+LYIPLK+VV  LSWQY
Sbjct: 299 LMIVVAAFLKYLNVVPKETQRGSKQLYKFISGNFTFPLMVGLGMLYIPLKDVVGMLSWQY 358

Query: 360 FIVVISVVLTVVSVGFFVSRFLNMSPVEAAIISACQSGMGGTGDVAILSTADRMNLMPFA 419
           F+VVISVV TV++ GFFVSRF+NM+PVEAAI+SACQSGMGGTGDVAILSTA+RM LMPFA
Sbjct: 359 FVVVISVVFTVIATGFFVSRFMNMNPVEAAIVSACQSGMGGTGDVAILSTANRMTLMPFA 418

Query: 420 QVATRLGGAITVITMTAILRIIF 442
           QVATRLGGAITVITMTAI R++F
Sbjct: 419 QVATRLGGAITVITMTAIFRMLF 441
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 356/419 (84%), Positives = 385/419 (90%)

Query: 27  KIGSVPLPVYVCLALLILLAGFLQKLPVNMLGGFAVILTMGWFLGTIGASIPGFKNFGGP 86
           +IGSVPLPVY+ LA LIL+ G+LQ+LPVNMLGGFAVILT+GW LGTIGA+IPG K+FGGP
Sbjct: 24  RIGSVPLPVYLVLATLILVTGWLQQLPVNMLGGFAVILTLGWLLGTIGATIPGLKHFGGP 83
```

```
Query: 87  AILSLLVPSILVFFNLINKNVLESTNMLMKQANFLYFYIACLVSGSILGMNRKMLIQGLL 146
           AILSLLVPSILVFFNL+N NVLE+TN+LMKQANFLYFYIACLV GSILGMNRK+LIQGL
Sbjct: 84  AILSLLVPSILVFFNLLNPNVLEATNVLMKQANFLYFYIACLVCGSILGMNRKILIQGLF 143

Query: 147 RMIFPMLLGMVCAMMVGTFVGVILGLEWRHTLFYIVTPVLAGGIGEGILPLSLGYSSITG 206
           RMI PMLLGMVCAM VGT VGVILGL+W+HTLFY+VTPVLAGGIGEGILPLSLGYS+ITG
Sbjct: 144 RMIIPMLLGMVCAMGVGTLVGVILGLDWQHTLFYVVTPVLAGGIGEGILPLSLGYSAITG 203

Query: 207 VASEQLVAQLIPATIIGNFFAILCTALLNRLGEKKPHLSGQGQLVRLNKGEDMSDIIADH 266
           V SEQLVAQLIPATIIGNFFAILCTALLNR GEK P  SGQGQLV++   EDMSD + D+
Sbjct: 204 VGSEQLVAQLIPATIIGNFFAILCTALLNRFGEKHPSYSGQGQLVKIGHSEDMSDALKDN 263

Query: 267 SGPIDVKKMGGGVLTACSLFIFGHLLQQLTGFPGPVLMIVAAAILKYINVIPRETQNGAK 326
           SG +DVK MG GVLTACSLFI G LLQ LT FPGPVLMI+ AA LKY+NVIP+ETQNGAK
Sbjct: 264 SGALDVKLMGAGVLTACSLFIAGGLLQHLTDFPGPVLMIILAAFLKYLNVIPQETQNGAK 323

Query: 327 QLYKFISGNFTFPLMAGLGLLYIPLKDVVATLSIQYFIVVISVVFTVISVGFFVSRFLNM 386
           QLYKFIS NFTFPLMAGLGLLYIPLK+VVATLS QYFIVVISVV TV+SVGFFVSRFLNM
Sbjct: 324 QLYKFISSNFTFPLMAGLGLLYIPLKEVVATLSWQYFIVVISVVLTVVSVGFFVSRFLNM 383

Query: 387 NPVEAGIISACQSGMGGTGDVAILSTADRMNLMPFAQVATRLGGAITVITMTAILRMLF 445
           +PVEA IISACQSGMGGTGDVAILSTADRMNLMPFAQVATRLGGAITVITMTAILR++F
Sbjct: 384 SPVEAAIISACQSGMGGTGDVAILSTADRMNLMPFAQVATRLGGAITVITMTAILRIIF 442
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 2248**

A DNA sequence (GBSx2369) was identified in *S.agalactiae* <SEQ ID 6953> which encodes the amino acid sequence <SEQ ID 6954>. This protein is predicted to be malic enzyme (mae). Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.28    Transmembrane  164 - 180 ( 164 - 181)

----- Final Results -----
               bacterial membrane --- Certainty=0.1914(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAB07709 GB:U35659 malic enzyme [Streptococcus bovis]
 Identities = 285/386 (73%), Positives = 332/386 (85%), Gaps = 1/386 (0%)

Query: 2   SENLGQLAINQARENGGKLEVISKVKVEDKRDLSIAYTPGVASVSSAIAEDVELAYELTT 61
           ++++ +LAI QA++ GGKLEV  KV +E K DL IAYTPGVA+VSSAI E  E AYELTT
Sbjct: 3   TKDVKELAIEQAKKFGGKLEVCPKVPIETKADLGIAYTPGVAAVSSAIYEKKERAYELTT 62

Query: 62  KKNTVAVVSDGSAVLGLGDIGPEAAMPVMEGKAALFKRFANVDAVPIVLKTNDTEEIISI 121
           KKNTVAV+SDGSAVLGLG+IGPEAAMPVMEGKAALFKRFA VD++P+VL T DTEEII
Sbjct: 63  KKNTVAVISDGSAVLGLGNIGPEAAMPVMEGKAALFKRFAGVDSIPLVLDTQDTEEIIQT 122

Query: 122 VKAISPTFGGINLEDISAPRCFEIEQRLIEECDIPVFHDDQHGTAIVVLAALFNSLKLVK 181
           VK ++PTFGGINLEDISAPRCFEIEQRLI+E DIPVFHDDQHGTAIVVLAAL+NSLKL+
Sbjct: 123 VKFLAPTFGGINLEDISAPRCFEIEQRLIDELDIPVFHDDQHGTAIVVLAALYNSLKLIN 182

Query: 182 KDIEDIRVVVNGGGSAGLSITRKLLSAGAKHVTVVDRFGIINDKDRESLAPHHKAIAKLT 241
           K IEDI VV+NGGGSAGLSITRK L+AG KH+ +VDR GI+++ D  +L PHH  IAKLT
Sbjct: 183 KKIEDIHVVINGGGSAGLSITRKFLAAGVKHIIIVDRTGILSETD-TALPPHHAEIAKLT 241

Query: 242 NREFQSGSLEDALENADVFIGVSAPEALHAEWISKMADKPIVFAMANPIPEIYPDQALKA 301
           NRE ++G L  ALE ADVF+GVSAP  L  EWI +M ++P++FAMANP+PEI+PD+AL A
```

```
Sbjct: 242 NREHRTGDLATALEGADVFVGVSAPGVLKPEWIQQMNEQPVIFAMANPVPEIFPDEALAA 301

Query: 302 GAYIVGTGRSDFPNQINNVLAFPGIFRGALDARAKTITVEMQIAAARGIASLIPEEELST 361
           GAYIVGTGRSDFPNQINNVLAFPGIFRGALDARAK IT+EMQIAAA+GIA LIP+ EL+
Sbjct: 302 GAYIVGTGRSDFPNQINNVLAFPGIFRGALDARAKKITIEMQIAAAKGIAKLIPDNELTP 361

Query: 362 THIIPNAFQNDVADVVAKSVSNAVQK 387
           T+IIP+ FQ  VA VVA+SV NAV++
Sbjct: 362 TNIIPDPFQEGVAKVVAESVRNAVKE 387
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6955> which encodes the amino acid sequence <SEQ ID 6956>. Analysis of this protein sequence reveals the following:

```
     Possible site: 48
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.44   Transmembrane  164 - 180 ( 164 - 181)
   INTEGRAL    Likelihood = -1.75   Transmembrane   94 - 110 (  94 - 110)

----- Final Results -----
               bacterial membrane --- Certainty=0.1977(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAB07709 GB:U35659 malic enzyme [Streptococcus bovis]
 Identities = 289/379 (76%), Positives = 334/379 (87%), Gaps = 1/379 (0%)

Query: 7   QLALEQAKTFGGKLEVQPKVDIKTKHDLSIAYTPGVASVSSAIAKDKTLAYDLTTKKNTV 66
           +LA+EQAK FGGKLEV PKV I+TK DL IAYTPGVA+VSSAI + K  AY+LTTKKNTV
Sbjct: 8   ELAIEQAKKFGGKLEVCPKVPIETKADLGIAYTPGVAAVSSAIYEKKERAYELTTKKNTV 67

Query: 67  AVISDGTAVLGLGDIGPEAAMPVMEGKAALFKAFAGVDAIPIVLDTKDTEEIISIVKALA 126
           AVISDG+AVLGLG+IGPEAAMPVMEGKAALFK FAGVD+IP+VLDT+DTEEII  VK LA
Sbjct: 68  AVISDGSAVLGLGNIGPEAAMPVMEGKAALFKRFAGVDSIPLVLDTQDTEEIIQTVKFLA 127

Query: 127 PTFGGINLEDISAPRCFEIEQRLIKECHIPVFHDDQHGTAIVVLAAIFNSLKLLKKSLDE 186
           PTFGGINLEDISAPRCFEIEQRLI E  IPVFHDDQHGTAIVVLAA++NSLKL+ K +++
Sbjct: 128 PTFGGINLEDISAPRCFEIEQRLIDELDIPVFHDDQHGTAIVVLAALYNSLKLINKKIED 187

Query: 187 VSIVVNGGGSAGLSITRKLLAAGATKVTVVDKFGIINEQEAAQLAPHHLDIAKVTNREFK 246
           + +V+NGGGSAGLSITRK LAAG   + +VD+ GI++E + A L PHH +IAK+TNRE +
Sbjct: 188 IHVVINGGGSAGLSITRKFLAAGVKHIIIVDRTGILSETDTA-LPPHHAEIAKLTNREHR 246

Query: 247 SGTLEDALEGADIFIGVSAPGVLKAEWISKMAARPVIFAMANPIPEIYPDEALEAGAYIV 306
           +G L  ALEGAD+F+GVSAPGVLK EWI +M  +PVIFAMANP+PEI+PDEAL AGAYIV
Sbjct: 247 TGDLATALEGADVFVGVSAPGVLKPEWIQQMNEQPVIFAMANPVPEIFPDEALAAGAYIV 306

Query: 307 GTGRSDFPNQINNVLAFPGIFRGALDARAKTITVEMQIAAAKGIASLVPDDALSTTNIIP 366
           GTGRSDFPNQINNVLAFPGIFRGALDARAK IT+EMQIAAAKGIA L+PD+ L+ TNIIP
Sbjct: 307 GTGRSDFPNQINNVLAFPGIFRGALDARAKKITIEMQIAAAKGIAKLIPDNELTPTNIIP 366

Query: 367 DAFKEGVAEIVAKSVRSVV 385
           D F+EGVA++VA+SVR+ V
Sbjct: 367 DPFQEGVAKVVAESVRNAV 385
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 306/387 (79%), Positives = 349/387 (90%)

Query: 1   MSENLGQLAINQARENGGKLEVISKVKVEDKRDLSIAYTPGVASVSSAIAEDVELAYELT 60
           M   LGQLA+ QA+  GGKLEV  KV ++ K DLSIAYTPGVASVSSAIA+D  LAY+LT
Sbjct: 1   MKNQLGQLALEQAKTFGGKLEVQPKVDIKTKHDLSIAYTPGVASVSSAIAKDKTLAYDLT 60

Query: 61  TKKNTVAVVSDGSAVLGLGDIGPEAAMPVMEGKAALFKRFANVDAVPIVLKTNDTEEIIS 120
           TKKNTVAV+SDG+AVLGLGDIGPEAAMPVMEGKAALFK FA VDA+PIVL T DTEEIIS
Sbjct: 61  TKKNTVAVISDGTAVLGLGDIGPEAAMPVMEGKAALFKAFAGVDAIPIVLDTKDTEEIIS 120
```

```
Query: 121 IVKAISPTFGGINLEDISAPRCFEIEQRLIEECDIPVFHDDQHGTAIVVLAALFNSLKLV 180
           IVKA++PTFGGINLEDISAPRCFEIEQRLI+EC IPVFHDDQHGTAIVVLAA+FNSLKL+
Sbjct: 121 IVKALAPTFGGINLEDISAPRCFEIEQRLIKECHIPVFHDDQHGTAIVVLAAIFNSLKLL 180

Query: 181 KKDIEDIRVVVNGGGSAGLSITRKLLSAGAKHVTVVDRFGIINDKDRESLAPHHKAIAKL 240
           KK ++++ +VVNGGGSAGLSITRKLL+AGA  VTVVD+FGIIN+++   LAPHH  IAK+
Sbjct: 181 KKSLDEVSIVVNGGGSAGLSITRKLLAAGATKVTVVDKFGIINEQEAAQLAPHHLDIAKV 240

Query: 241 TNREFQSGSLEDALENADVFIGVSAPEALHAEWISKMADKPIVFAMANPIPEIYPDQALK 300
           TNREF+SG+LEDALE AD+FIGVSAP  L AEWISKMA +P++FAMANPIPEIYPD+AL+
Sbjct: 241 TNREFKSGTLEDALEGADIFIGVSAPGVLKAEWISKMAARPVIFAMANPIPEIYPDEALE 300

Query: 301 AGAYIVGTGRSDFPNQINNVLAFPGIFRGALDARAKTITVEMQIAAARGIASLIPEEELS 360
           AGAYIVGTGRSDFPNQINNVLAFPGIFRGALDARAKTITVEMQIAAA+GIASL+P++ LS
Sbjct: 301 AGAYIVGTGRSDFPNQINNVLAFPGIFRGALDARAKTITVEMQIAAAKGIASLVPDDALS 360

Query: 361 TTHIIPNAFQNDVADVVAKSVSNAVQK 387
           TT+IIP+AF+  VA++VAKSV + V K
Sbjct: 361 TTNIIPDAFKEGVAEIVAKSVRSVVLK 387
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 2249**

A DNA sequence (GBSx2370) was identified in *S.agalactiae* <SEQ ID 6957> which encodes the amino acid sequence <SEQ ID 6958>. This protein is predicted to be Bta. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.02   Transmembrane   29 -  45 (  29 -  45)

----- Final Results -----
               bacterial membrane --- Certainty=0.1808(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD56628 GB:AF165218 Bta [Streptococcus pneumoniae]
 Identities = 35/112 (31%), Positives = 63/112 (56%)

Query: 1   MYSFEELLATMTLITAAEIEDKIDSNQDFVLFIGRISCPFCHLFVPKIVEVADEDEFELF 60
           M  F + +  + + T    ++ +D  +    FIGR +CP+C  F   +  V  E +  ++
Sbjct: 1   MEQFLDNIKDLEVTTVVRAQEALDKKETATFFIGRKTCPYCRKFAGTLSGVVAETKAHIY 60

Query: 61  HLDSEDFDHWTANKEFRNKYDIPTVPGLMVVKNGTIKVKCDSKMTKEEIREF 112
            ++SE+       + FR++Y IPTVPG + + +G I V+CDS M+ +EI++F
Sbjct: 61  FINSEEASQLNDLQAFRSRYGIPTVPGFVHITDGQINVRCDSSMSAQEIKDF 112
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6959> which encodes the amino acid sequence <SEQ ID 6960>. Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0900(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 39/111 (35%), Positives = 66/111 (59%)

Query: 3   SFEELLATMTLITAAEIEDKIDSNQDFVLFIGRISCPFCHLFVPKIVEVADEDEFELFHL 62
            +FEE++A     + AE+   I S +D ++F+GR SCP+C  F PK+ +VA +++ E++ +
Sbjct: 11  TFEEIVANFIPSSVAEVTSAIASGKDMIVFLGRSSCPYCRRFAPKLAQVATDNQKEVYFV 70

Query: 63  DSEDFDHWTANKEFRNKYDIPTVPGLMVVKNGTIKVKCDSKMTKEEIREFI 113
            DSE+         FR  Y + TVP L+V  +   +  CDS +T ++I  F+
Sbjct: 71  DSENAADAAELAAFRENYQLVTVPALLVSYDQHQRAVCDSSLTPDDILAFL 121
```

SEQ ID 6958 (GBS427) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 80 (lane 5; MW 16.2kDa).

GBS427-His was purified as shown in Figure 214, lane 8.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

## Example 2250

A DNA sequence (GBSx2371) was identified in *S.agalactiae* <SEQ ID 6961> which encodes the amino acid sequence <SEQ ID 6962>. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -7.75    Transmembrane    2 -  18 (   1 -  21)

----- Final Results -----
               bacterial membrane --- Certainty=0.4100(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9437> which encodes amino acid sequence <SEQ ID 9438> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAA11328 GB:D78257 ORF11 [Enterococcus faecalis]
 Identities = 36/80 (45%), Positives = 58/80 (72%)

Query: 1  MSLPIIMLVVMVGMMFFMQRQQKKQAQERQKQLNAVQKGDEIVTIGGLFGVVDEVNTEAQ 60
           M L +IML+V+V M F++ R QKKQ +ERQ  LN +Q GD +VTIGGL GV+ E++++ +
Sbjct: 1  MKLMLIMLLVIVAMYFYLFRTQKKQQKERQDFLNNLQPGDAVVTIGGLHGVISEISSDKK 60

Query: 61 RMVLDVDGVYLTFELAAIKS 80
           ++ LD +G +  F+  +I++
Sbjct: 61 KVTLDCEGAFFDFDQQSIRT 80
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6963> which encodes the amino acid sequence <SEQ ID 6964>. Analysis of this protein sequence reveals the following:

```
     Possible site: 60
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -6.10    Transmembrane    3 -  19 (   1 -  22)
   INTEGRAL    Likelihood = -3.03    Transmembrane   63 -  79 (  63 -  79)

----- Final Results -----
               bacterial membrane --- Certainty=0.3442(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:BAA11328 GB:D78257 ORF11 [Enterococcus faecalis]
 Identities = 29/75 (38%), Positives = 52/75 (68%)

Query: 6  ILMFVVMLGLIWFMQRQQKKQAQERQNQLNAIEKGDEVVTIGGMFAIVDEVDTTAKKIVL 65
          ++M +V++ + +++ R QKKQ +ERQ+ LN ++ GD VVTIGG+  ++ E+ +  KK+ L
Sbjct: 5  LIMLLVIVAMYFYLFRTQKKQQKERQDFLNNLQPGDAVVTIGGLHGVISEISSDKKKVTL 64

Query: 66 DVDGVFLTFELLAIK 80
          D +G F  F+  +I+
Sbjct: 65 DCEGAFFDFDQQSIR 79
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 63/90 (70%), Positives = 80/90 (88%)

Query: 4  PIIMLVVMVGMMFFMQRQQKKQAQERQKQLNAVQKGDEIVTIGGLFGVVDEVNTEAQRMV 63
          PI+M VVM+G+++FMQRQQKKQAQERQ QLNA++KGDE+VTIGG+F +VDEV+T A+++V
Sbjct: 5  PILMFVVMLGLIWFMQRQQKKQAQERQNQLNAIEKGDEVVTIGGMFAIVDEVDTTAKKIV 64

Query: 64 LDVDGVYLTFELAAIKSVVSKAATPTEPVE 93
          LDVDGV+LTFEL AIK +V+KA T T  VE
Sbjct: 65 LDVDGVFLTFELLAIKRIVTKATTETTLVE 94
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## Example 2251

A DNA sequence (GBSx2372) was identified in *S.agalactiae* <SEQ ID 6965> which encodes the amino acid sequence <SEQ ID 6966>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

## Example 2252

A DNA sequence (GBSx2373) was identified in *S.agalactiae* <SEQ ID 6967> which encodes the amino acid sequence <SEQ ID 6968>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.38   Transmembrane  164 - 180 ( 164 - 180)

----- Final Results -----
               bacterial membrane --- Certainty=0.1553(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB61731 GB:AL133220 putative oxidoreductase. [Streptomyces
```

```
          coelicolor A3(2)]
 Identities = 72/216 (33%), Positives = 120/216 (55%), Gaps = 1/216 (0%)

Query: 14  AQALEARGQKLYSVANRTYDKGLEFATKYGIQKVYDHIDQVFEDPEVDIIYISTPHNTHI 73
            A  ++    ++ +VA+RT      FA ++GI + Y   + +  D +VD++Y++TPH+ H
Sbjct: 25  ADLVDLPDAEVVAVASRTEASAKTFAERFGIPRAYGGWETLARDEDVDVVYVATPHSAHR 84

Query: 74  SFLRKALANGKHVLCEKSITLNSTELKEAIDLAETNHVVLAEAMTIFHMPIYRQLKTLVD 133
            +     L  G++VLCEK  TLN+ E  E + LA  N V L EAM ++  P+ R+LK LV
Sbjct: 85  TAAGLCLEAGRNVLCEKPFTLNAREAAELVALARENGVFLMEAMWMYCNPLVRRLKELVA 144

Query: 134 SGKLGPLKMIQMNFGSYKEYDMTNRFFSRDLAGGALLDIGVYALSCIRWFMSEAPHNITS 193
             G +G ++ +Q +FG    +   +R       GGALLD+GVY +S  +  + E P ++ +
Sbjct: 145 DGAIGEVRSLQADFGLAGPFPAAHRLRDPAQGGGALLDLGVYPVSFAQLLLGE-PTDVAA 203

Query: 194 QVTFAPTGVDEQVGILLTNPANEMATVSLSLHAKQP 229
            +   +  GVD Q G LL+   + +A++  S+    P
Sbjct: 204 RAVLSEEGVDLQTGALLSYGNDALASIHCSITGGTP 239
```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2253**

A DNA sequence (GBSx2374) was identified in *S.agalactiae* <SEQ ID 6969> which encodes the amino acid sequence <SEQ ID 6970>. This protein is predicted to be surface protein Rib. Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4957(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2254**

A DNA sequence (GBSx2375) was identified in *S.agalactiae* <SEQ ID 6971> which encodes the amino acid sequence <SEQ ID 6972>. This protein is predicted to be surface protein Rib. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1892(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2255**

A DNA sequence (GBSx2376) was identified in *S.agalactiae* <SEQ ID 6973> which encodes the amino acid sequence <SEQ ID 6974>. This protein is predicted to be a host cell surface-exposed lipoprotein. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -7.75    Transmembrane    9 -  25 (   5 -  28)

----- Final Results -----
               bacterial membrane --- Certainty=0.4100(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9005> which encodes amino acid sequence <SEQ ID 9006> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 3
SRCFLG: 0
McG: Length  of  UR:   24
     Peak Value of UR:   2.84
     Net Charge of CR: 2
McG: Discrim Score:     10.29
GvH: Signal Score (-7.5): -4.34
     Possible site: 34
>>> Seems to have an uncleavable N-term signal seq
Amino Acid Composition: calculated from 1
ALOM program   count: 1 value:  -7.75 threshold:  0.0
   INTEGRAL    Likelihood = -7.75    Transmembrane    5 -  21 (   1 -  24)
   PERIPHERAL  Likelihood = 13.31      86
 modified ALOM score:   2.05
icml HYPID: 7  CFP: 0.410

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.4100(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC03455 GB:AF020798 putative host cell surface-exposed
           lipoprotein [Streptococcus thermophilus bacteriophage TP-J34]
 Identities = 40/102 (39%), Positives = 63/102 (61%), Gaps = 10/102 (9%)

Query: 101 KNALISAKIYSKTMNLSKQSIFEQLYSESPDKATHSDKFTKEESQYAIDHLKVDFKENAL 160
           + A+  AK Y+ T+++SK+ +  QL S         DK++++ S YA+++  +D+ + AL
Sbjct: 51  RTAVSKAKQYASTVHMSKEELRSQLVS--------FDKYSQDASDYAVENSGIDYNKQAL 102

Query: 161 ETAKSYQSSSSLSKEEIYKQLTSTLGDKFTNDEAQYAVDHLK 202
           E AK YQ + S+S + I  QL S   DKFT +EA YAV +LK
Sbjct: 103 EKAKQYQDTLSMSPDAIRDQLVSF--DKFTQEEADYAVANLK 142
 Identities = 40/112 (35%), Positives = 64/112 (56%), Gaps = 9/112 (8%)

Query: 41  KKAKIKFNKTQKKIVKKAREYAKSGHMSKDSIIEKLKKDSKKYRQEDINFVINNLKVDYK 100
           + ++ K  K  +  V KA++YA + HMSK+ +  +L    K Y Q+  ++ + N  +DY
Sbjct: 40  QSSESKVPKEYRTAVSKAKQYASTVHMSKEELRSQLVSFDK-YSQDASDYAVENSGIDYN 98

Query: 101 KNALISAKIYSKTMNLSKQSIFEQLYSESPDKATHSDKFTKEESQYAIDHLK 152
           K AL  AK Y  T+++S  +I +QL S         DKFT+EE+ YA+ +LK
Sbjct: 99  KQALEKAKQYQDTLSMSPDAIRDQLVS--------FDKFTQEEADYAVANLK 142
```

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 9006 (GBS122) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 38 (lane 6; MW 21.9kDa).

GBS122-His was purified as shown in Figure 202, lane 8.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2256

A DNA sequence (GBSx2377) was identified in *S.agalactiae* <SEQ ID 6975> which encodes the amino acid sequence <SEQ ID 6976>. This protein is predicted to be transposase (orfA). Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2830(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB90833 GB:AJ250837 hypothetical protein [Streptococcus dysgalactiae]
 Identities = 91/96 (94%), Positives = 93/96 (96%)

Query: 1  MSRKVRRHFTDDFKQQIVDLYNVGRKRSSLIKVYELTPSTFDKWVRQAKTTGSFKSIDNL 60
          MSRK+RRHFTDDFKQQIVDLYN GRKRSSLIK YELTPSTFDKWVRQAKTTGSFKS+DNL
Sbjct: 1  MSRKIRRHFTDDFKQQIVDLYNAGRKRSSLIKEYELTPSTFDKWVRQAKTTGSFKSVDNL 60

Query: 61 TDEQRELIELRKHNKELEMQLDILKQAAVIMAQKGK 96
          TDEQRELIELRK NKELEMQLDILKQAAVIMAQKGK
Sbjct: 61 TDEQRELIELRKRNKELEMQLDILKQAAVIMAQKGK 96
```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2257

A DNA sequence (GBSx2378) was identified in *S.agalactiae* <SEQ ID 6977> which encodes the amino acid sequence <SEQ ID 6978>. This protein is predicted to be transposase (orfB). Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2618(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9915> which encodes amino acid sequence <SEQ ID 9916> was also identified.

A related GBS nucleic acid sequence <SEQ ID 9903> which encodes amino acid sequence <SEQ ID 9904> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB90834 GB:AJ250837 putative transposase [Streptococcus dysgalactiae]
 Identities = 243/259 (93%), Positives = 250/259 (95%)

Query: 1   MCRWLNMPHSSYYYQAVESVSETEFEETIKRIFLDSESRYGSRKIKICLNNEGITLSRRR 60
           MCRWLN+P SSYYY+AVE VSE E EE+IK IFL+S++RYGSRKIKICLNNEGITLSRRR
Sbjct: 1   MCRWLNIPRSSYYYKAVEPVSEAELEESIKAIFLESKARYGSRKIKICLNNEGITLSRRR 60

Query: 61  IRRIMKRLNLVSVYQKATFKPHSRGKNEAPIPNHLDRQFKQERPLQALVTDLTYVRVGNR 120
           IRRIMKRLNLVSVYQKATFKPHSRGKNEAPIPNHLDRQFK ERPLQALVTDLTYVRVGNR
Sbjct: 61  IRRIMKRLNLVSVYQKATFKPHSRGKNEAPIPNHLDRQFKPERPLQALVTDLTYVRVGNR 120

Query: 121 WAYVCLIIDLYNREIIGLSLGWHKTAELVKQAIQSIPYALTKVKMFHSDRGKEFDNQLID 180
           WAYVCLIIDLYNREIIGLSLGWHKTAELVKQAIQSIPY LTKVKMFHSDRGKEF+NQLID
Sbjct: 121 WAYVCLIIDLYNREIIGLSLGWHKTAELVKQAIQSIPYPLTKVKMFHSDRGKEFNNQLID 180

Query: 181 EILEAFGITRSLSQAGCPYDNAVAESTYRAFKIEFVYQETFQLLEELALKTKDYVHWWNY 240
           EILEAFGITRSLSQAGCPYDNAVAESTYRAFKIEFVYQETFQ LEELALKTK YVHWWNY
Sbjct: 181 EILEAFGITRSLSQAGCPYDNAVAESTYRAFKIEFVYQETFQSLEELALKTKAYVHWWNY 240

Query: 241 HRIHGSLNYQTPMTKRLIA 259
           HRIHGSLNYQTPMTKRLIA
Sbjct: 241 HRIHGSLNYQTPMTKRLIA 259
```

There is also homology to SEQ ID 32.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2258**

A DNA sequence (GBSx2379) was identified in *S.agalactiae* <SEQ ID 6979> which encodes the amino acid sequence <SEQ ID 6980>. This protein is predicted to be pXO1-128. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3684(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD32432 GB:AF065404 pXO1-128 [Bacillus anthracis]
 Identities = 45/69 (65%), Positives = 52/69 (75%)

Query: 17 MKKAGKSNRVIMETLGIKNNSQIYTWMKWYENEELYRFHQGVGKQYTYGKGLEHLSEVEQ 76
          MKK   SNR IME LGIKN SQI TWMKWY  ++ YRF Q VGKQY+YGKG + LSE+EQ
Sbjct: 1  MKKESYSNRTIMEKLGIKNVSQIKTWMKWYRTDQTYRFQQPVGKQYSYGKGPKELSELEQ 60

Query: 77 LQLQVDLLK 85
          L+L+   LK
Sbjct: 61 LRLENKHLK 69
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2259**

A DNA sequence (GBSx2380) was identified in *S.agalactiae* <SEQ ID 6981> which encodes the amino acid sequence <SEQ ID 6982>. This protein is predicted to be transposase. Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes.*

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2260**

A DNA sequence (GBSx2382) was identified in *S.agalactiae* <SEQ ID 6985> which encodes the amino acid sequence <SEQ ID 6986>. This protein is predicted to be Lmb. Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> May be a lipoprotein

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1595> which encodes the amino acid sequence <SEQ ID 1596>. Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> May be a lipoprotein

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 302/306 (98%), Positives = 303/306 (98%)

Query: 1   MKKVFFLMAMVVSLVMIAGCDKSANPKQPTQGMSVVTSFYPMYAMTKEVSGDLNDVRMIQ 60
            MKK FFLMAMVVSLVMIAGCDKSANPKQPTQGMSVVTSFYPMYAMTKEVSGDLNDVRMIQ
Sbjct: 1   MKKGFFLMAMVVSLVMIAGCDKSANPKQPTQGMSVVTSFYPMYAMTKEVSGDLNDVRMIQ 60

Query: 61  SGAGIHSFEPSVNDVAAIYDADLFVYHSHTLEAWARDLDPNLKKSKVNVFEASKPLTLDR 120
            SGAGIHSFEPSVNDVAAIYDADLFVYHSHTLEAWARDLDPNLKKSKV+VFEASKPLTLDR
Sbjct: 61  SGAGIHSFEPSVNDVAAIYDADLFVYHSHTLEAWARDLDPNLKKSKVDVFEASKPLTLDR 120

Query: 121 VKGLEDMEVTQGIDPATLYDPHTWTDPVLAGEEAVNIAKELGHLDPKHKDSYTKKAKAFK 180
            VKGLEDMEVTQGIDPATLYDPHTWTDPVLAGEEAVNIAKELG LDPKHKDSYTK AKAFK
Sbjct: 121 VKGLEDMEVTQGIDPATLYDPHTWTDPVLAGEEAVNIAKELGRLDPKHKDSYTKNAKAFK 180

Query: 181 KEAEQLTEEYTQKFKKVRSKTFVTQHTAFSYLAKRFGLKQLGISGISPEQEPSPRQLKEI 240
```

```
           KEAEQLTEEYTQKFKKVRSKTFVTQHTAFSYLAKRFGLKQLGISGISPEQEPSPRQLKEI
Sbjct: 181 KEAEQLTEEYTQKFKKVRSKTFVTQHTAFSYLAKRFGLKQLGISGISPEQEPSPRQLKEI 240

Query: 241 QDFVKEYNVKTIFAEDNVNPKIAHAIAKSTGAKVKTLSPLEAAPSGNKTYLENLRANLEV 300
           QDFVKEYNVKTIFAEDNVNPKIAHAIAKSTGAKVKTLSPLEAAPSGNKTYLENLRANLEV
Sbjct: 241 QDFVKEYNVKTIFAEDNVNPKIAHAIAKSTGAKVKTLSPLEAAPSGNKTYLENLRANLEV 300

Query: 301 LYQQLK 306
           LYQQLK
Sbjct: 301 LYQQLK 306
```

There is also homology to SEQ ID 4.

SEQ ID 6986 (GBS189) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 38 (lane 2; MW 35.2kDa).

The GBS189-His fusion product was purified (Figure 204, lane 7) and used to immunise mice. The resulting antiserum was used for Western blot (Figure 248A), FACS (Figure 248B), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 2261**

A DNA sequence (GBSx2383) was identified in *S.agalactiae* <SEQ ID 6987> which encodes the amino acid sequence <SEQ ID 6988>. Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4656(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAB41455 GB:U34956 phosphoribosylformylglycinamidine synthase
           [Mycobacterium tuberculosis]
 Identities = 73/237 (30%), Positives = 112/237 (46%), Gaps = 25/237 (10%)

Query: 43  GAGGVCVAIGELAD----GLEIDLDKVPLKYQGLNGTEIAISESQERMSVVVGPSDVDAF 98
           G  G+  A  ELA     G+ I LD VPL+ + +   E+  SESQERM  VV P +VDAF
Sbjct: 282 GGAGLSCATSELASAGDGGMTIQLDSVPLRAKEMTPAEVLCSESQERMCAVVSPKNVDAF 341

Query: 99  IAACNKENIDAVVVATVTEKPNLVMTWNGETIVDLERCFLDTNG------VRVVVDAKVV 152
           +A C K  + A V+  VT+   L +TW+GET+VD+    +   G      V        +
Sbjct: 342 LAVCRKWEVLATVIGEVTDGDRLQITWHGETVVDVPPRTVAHEGPVYQRPVARPDTQDAL 401

Query: 153 DKDLTVPEARTTSAETLEADMLKVLSDLNHASQKGLQTIFDSSVGRSTV--NHPIGGRYQ 210
           + D +   +R  + + L A +L +L   +  S+  +   +D  V  +TV   H  GG  +
Sbjct: 402 NADRSAKLSRPVTGDELRATLLALLGSPHLCSRAFITEQYDRYVRGNTVLAEHADGGMLR 461

Query: 211 ITPTESSVQKLPVQYGVTTTASVMAQGYNPYIAEWSPYHGAAYAVIEATARLVATGA 267
           I   ES+ + + V    +   +++            PY GA  A+ EA   +  TGA
Sbjct: 462 I--DESTGRGIAVSTDASGRYTLL-----------DPYAGAQLALAEAYRNVAVTGA 505
```

There is also homology to SEQ ID 982.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2262**

A DNA sequence (GBSx2384) was identified in *S.agalactiae* <SEQ ID 6989> which encodes the amino acid sequence <SEQ ID 6990>. This protein is predicted to be 30S ribosomal protein S11 (rpsK). Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0598(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9281> which encodes amino acid sequence <SEQ ID 9282> was also identified. A further related GBS nucleic acid sequence <SEQ ID 10919> which encodes amino acid sequence <SEQ ID 10920> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB11918 GB:Z99104 ribosomal protein S11 (BS11) [Bacillus subtilis]
 Identities = 81/92 (88%), Positives = 87/92 (94%)

Query: 2   HGNALAWSSAGALGFKGSRKSTPFAAQMAAEAAAKSAQEHGLKTVEVTVKGPGSGRESAI 61
            HGNA++WSSAGALGF+GSRKSTPFAAQMAAE AAK + EHGLKT+EVTVKGPGSGRE+AI
Sbjct: 40  HGNAISWSSAGALGFRGSRKSTPFAAQMAAETAAKGSIEHGLKTLEVTVKGPGSGREAAI 99

Query: 62  RALAAAGLEVTAIRDVTPVPHNGARPPKRRRV 93
            RAL AAGLEVTAIRDVTPVPHNG RPPKRRRV
Sbjct: 100 RALQAAGLEVTAIRDVTPVPHNGCRPPKRRRV 131
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6991> which encodes the amino acid sequence <SEQ ID 6992>. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0945(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 92/93 (98%), Positives = 93/93 (99%)

Query: 1   MHGNALAWSSAGALGFKGSRKSTPFAAQMAAEAAAKSAQEHGLKTVEVTVKGPGSGRESA 60
           +HGNALAWSSAGALGFKGSRKSTPFAAQMAAEAAAKSAQEHGLKTVEVTVKGPGSGRESA
Sbjct: 35  VHGNALAWSSAGALGFKGSRKSTPFAAQMAAEAAAKSAQEHGLKTVEVTVKGPGSGRESA 94

Query: 61  IRALAAAGLEVTAIRDVTPVPHNGARPPKRRRV 93
           IRALAAAGLEVTAIRDVTPVPHNGARPPKRRRV
Sbjct: 95  IRALAAAGLEVTAIRDVTPVPHNGARPPKRRRV 127
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## Example 2263

A DNA sequence (GBSx2385) was identified in *S.agalactiae* <SEQ ID 6993> which encodes the amino acid sequence <SEQ ID 6994>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2551(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB03881 GB:AP001507 DNA-directed RNA polymerase alpha subunit
          [Bacillus halodurans]
 Identities = 190/314 (60%), Positives = 249/314 (78%), Gaps = 2/314 (0%)

Query: 1   MIEFEKPIITKIDENKD--YGRFVIEPLERGYGTTLGNSLRRVLLSSLPGAAVTSIKIDG 58
           MIE EKP+I  I+ ++D  YG+FV+EPLERGYGTTLGNSLRR+LLSSLPGAAVTS++IDG
Sbjct: 1   MIEIEKPVIETIEISEDAKYGKFVVEPLERGYGTTLGNSLRRILLSSLPGAAVTSVQIDG 60

Query: 59  VLHEFDTIPGVREDVMQIILNVKGLAVKSYVEDEKIIELDVEGPAEITAGDILTDSDIEI 118
           VLHEF TI GV EDV  I+LN+K LA+K Y +++K +E+D +G   +TAGD+  DSD+++
Sbjct: 61  VLHEFSTIEGVVEDVTTIVLNLKQLALKIYSDEDKTLEIDTQGEGVVTAGDLTHDSDVDV 120

Query: 119 VNPDHYLFTIAEGHSLKATMTVAKNRGYVPAEGNKKDDAPVGTLAVDSIYTPVKKVNYQV 178
           +NPD ++ T+  G  L+  +T  + RGYVPAEGNK D+  +G + +DSIYTPV +VNYQV
Sbjct: 121 LNPDLHIATLTTGAHLRMRITAKRGRGYVPAEGNKSDELAIGVIPIDSIYTPVSRVNYQV 180

Query: 179 EPARVGSNDGFDKLTIEIMTNGTIIPEDALGLSARVLIEHLNLFTDLTEVAKATEVMKET 238
           E  RVG    +DKLT+++ T+G+I PE+A+ L A++L EHLN+F  LT+ A+  E+M E
Sbjct: 181 ENTRVGQVTNYDKLTLDVWTDGSIRPEEAVSLGAKILTEHLNIFVGLTDQAQNAEIMVEK 240

Query: 239 EKVNDEKVLDRTIEELDLSVRSYNCLKRAGINTVFDLTEKTEPEMMKVRNLGRKSLEEVK 298
           E+   EKVL+ TIEELDLSVRSYNCLKRAGINTV +LT+KTE +MMKVRNLGRKSLEEV+
Sbjct: 241 EEDQKEKVLEMTIEELDLSVRSYNCLKRAGINTVQELTQKTEEDMMKVRNLGRKSLEEVQ 300

Query: 299 IKLADLGLGLKNDK 312
            KL +LGLGL+ ++
Sbjct: 301 EKLGELGLGLRKEE 314
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6995> which encodes the amino acid sequence <SEQ ID 6996>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2551(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 305/312 (97%), Positives = 311/312 (98%)

Query: 1   MIEFEKPIITKIDENKDYGRFVIEPLERGYGTTLGNSLRRVLLSSLPGAAVTSIKIDGVL 60
           MIEFEKPIITKIDENKDYGRFVIEPLERGYGTTLGNSLRRVLLSSLPGAAVTSIKIDGVL
Sbjct: 1   MIEFEKPIITKIDENKDYGRFVIEPLERGYGTTLGNSLRRVLLSSLPGAAVTSIKIDGVL 60

Query: 61  HEFDTIPGVREDVMQIILNVKGLAVKSYVEDEKIIELDVEGPAEITAGDILTDSDIEIVN 120
           HEFDTIPGVREDVMQIILNVKGLAVKSYVEDEKIIEL+VEGPAE+TAGDILTDSDIE+VN
Sbjct: 61  HEFDTIPGVREDVMQIILNVKGLAVKSYVEDEKIIELEVEGPAEVTAGDILTDSDIELVN 120
```

```
Query: 121 PDHYLFTIAEGHSLKATMTVAKNRGYVPAEGNKKDDAPVGTLAVDSIYTPVKKVNYQVEP 180
           PDHYLFTIAEGHSL+ATMTVAK RGYVPAEGNKKDDAPVGTLAVDSIYTPVKKVNYQVEP
Sbjct: 121 PDHYLFTIAEGHSLRATMTVAKKRGYVPAEGNKKDDAPVGTLAVDSIYTPVKKVNYQVEP 180

Query: 181 ARVGSNDGFDKLTIEIMTNGTIIPEDALGLSARVLIEHLNLFTDLTEVAKATEVMKETEK 240
           ARVGSNDGFDKLTIEIMTNGTIIPEDALGLSARVLIEHLNLFTDLTEVAKATEVMKETEK
Sbjct: 181 ARVGSNDGFDKLTIEIMTNGTIIPEDALGLSARVLIEHLNLFTDLTEVAKATEVMKETEK 240

Query: 241 VNDEKVLDRTIEELDLSVRSYNCLKRAGINTVFDLTEKTEPEMMKVRNLGRKSLEEVKIK 300
           VNDEKVLDRTIEELDLSVRSYNCLKRAGINTVFDLTEK+EPEMMKVRNLGRKSLEEVK+K
Sbjct: 241 VNDEKVLDRTIEELDLSVRSYNCLKRAGINTVFDLTEKSEPEMMKVRNLGRKSLEEVKVK 300

Query: 301 LADLGLGLKNDK 312
           LADLGLGLKNDK
Sbjct: 301 LADLGLGLKNDK 312
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 2264**

A DNA sequence (GBSx2386) was identified in *S.agalactiae* <SEQ ID 6997> which encodes the amino acid sequence <SEQ ID 6998>. This protein is predicted to be 50S ribosomal protein L17 (rplQ). Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1609(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB11920 GB:Z99104 ribosomal protein L17 (BL15) [Bacillus subtilis]
 Identities = 95/128 (74%), Positives = 105/128 (81%), Gaps = 8/128 (6%)

Query: 1   MAYRKLGRTSSQRKAMLRDLTTDLLINESIVTTEARAKEIRKTVEKMITLGKRGDLHARR 60
           M+YRKLGRTS+QRKAMLRDLTTDL+INE I TTE RAKE+R  VEKMITLGKRGDLHARR
Sbjct: 1   MSYRKLGRTSAQRKAMLRDLTTDLIINERIETTETRAKELRSVVEKMITLGKRGDLHARR 60

Query: 61  QAAAYVRNEIASENYDEASDKYTSTTALQKLFDDIAPRYAERNGGYTRILKTEPRRGDAA 120
           QAAAY+RNE+A+E  ++         ALQKLF DIA RY ER GGYTRI+K  PRRGD A
Sbjct: 61  QAAAYIRNEVANEENNQ--------DALQKLFSDIATRYEERQGGYTRIMKLGPRRGDGA 112

Query: 121 PMAIIELV 128
           PMAIIELV
Sbjct: 113 PMAIIELV 120
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6999> which encodes the amino acid sequence <SEQ ID 7000>. Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1609(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 125/128 (97%), Positives = 127/128 (98%)
```

```
Query: 1   MAYRKLGRTSSQRKAMLRDLTTDLLINESIVTTEARAKEIRKTVEKMITLGKRGDLHARR 60
           MAYRKLGRTSSQRKAMLRDLTTDLLINESIVTTEARAKEIRKTVEKMITLGKRGDLHARR
Sbjct: 1   MAYRKLGRTSSQRKAMLRDLTTDLLINESIVTTEARAKEIRKTVEKMITLGKRGDLHARR 60

Query: 61  QAAAYVRNEIASENYDEASDKYTSTTALQKLFDDIAPRYAERNGGYTRILKTEPRRGDAA 120
           QAAAYVRNEIASENYDEA+DKYTSTTALQKLF +IAPRYAERNGGYTRILKTEPRRGDAA
Sbjct: 61  QAAAYVRNEIASENYDEATDKYTSTTALQKLFSEIAPRYAERNGGYTRILKTEPRRGDAA 120

Query: 121 PMAIIELV 128
           PMAIIELV
Sbjct: 121 PMAIIELV 128
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 2265**

A DNA sequence (GBSx2396) was identified in *S.agalactiae* <SEQ ID 7001> which encodes the amino acid sequence <SEQ ID 7002>. This protein is predicted to be mercuric reductase. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2384(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAA83977 GB:AF138877 mercuric reductase MerA [Bacillus sp.
           RC607]
 Identities = 29/33 (87%), Positives = 32/33 (96%)

Query: 4   VGLTEEQAKEKGYDVKTSVLPLXAVPRAIVNRE 36
           VGLTE+QAKEKGY+VKTSVLPL AVPRA+VNRE
Sbjct: 520 VGLTEQQAKEKGYEVKTSVLPLDAVPRALVNRE 552
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2266**

A DNA sequence (GBSx2397) was identified in *S.agalactiae* <SEQ ID 7003> which encodes the amino acid sequence <SEQ ID 7004>. This protein is predicted to be mercuric reductase. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3016(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA70224 GB:Y09024 mercuric reductase [Bacillus cereus]
```

```
 Identities = 146/194 (75%), Positives = 175/194 (89%)

Query: 2   PQISGLEKMDYLTSTTLLELKKIPKRLTVIGSGYIGMELGQLFHHLGSEITLMQRSERLL 61
            P I GL ++DYLTST+LLELKK+PKRL VIGSGYIGMELGQLFH+LGSE+TL+QRSERLL
Sbjct: 226 PNIPGLNEVDYLTSTSLLELKKVPKRLVVIGSGYIGMELGQLFHNLGSEVTLIQRSERLL 285

Query: 62  KEYDPEISESVEKALIEQGINLVKGATFERVEQSGEIKRVYVTVNGSREVIESDQLLVAT 121
            KEYDPEISESVEK+L+EQGINLVKGAT+ER+EQ+G+IK+V+V VNG + +IE+DQLLVAT
Sbjct: 286 KEYDPEISESVEKSLVEQGINLVKGATYERIEQNGDIKKVHVEVNGKKRIIEADQLLVAT 345

Query: 122 GRKPNTDSLNLSAAGVETGKNNEILINDFGQTSNEKIYAAGDVTLGPQFVYVAAYEGGII 181
            GR PNT +LNL AAGVE G   EI+I+D+ +T+N +IYAAGDVTLGPQFVYVAAY+GG+
Sbjct: 346 GRTPNTATLNLRAAGVEIGSRGEIIIDDYSRTTNTRIYAAGDVTLGPQFVYVAAYQGGVA 405

Query: 182 TDNAIGGLNKKIDL 195
              NAIGGLNKK++L
Sbjct: 406 APNAIGGLNKKLNL 419
```

There is also homology to SEQ ID 1820.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2267**

A DNA sequence (GBSx2398) was identified in *S.agalactiae* <SEQ ID 7005> which encodes the amino acid sequence <SEQ ID 7006>. This protein is predicted to be triacylglycerol acylhydrolase. Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3180(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2268**

A DNA sequence (GBSx2399) was identified in *S.agalactiae* <SEQ ID 7007> which encodes the amino acid sequence <SEQ ID 7008>. Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0544(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC74453 GB:AE000234 orf, hypothetical protein [Escherichia
          coli K12]
 Identities = 45/58 (77%), Positives = 51/58 (87%)
```

```
Query: 1  MPWQNLLHAGQENLFSGLTALTAEFTVGEGKLMTHDEPCSMAPDDKHDLISGTCSHLP 58
          +PWQNLLHAG+ENLFSGLTAL+AEFT+GEG+LM HD P    APD+  DLISGTCSHLP
Sbjct: 34 LPWQNLLHAGEENLFSGLTALSAEFTIGEGELMAHDVPLGCAPDEYDDLISGTCSHLP 91
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2269**

A DNA sequence (GBSx2400) was identified in *S.agalactiae* <SEQ ID 7009> which encodes the amino acid sequence <SEQ ID 7010>. This protein is predicted to be transposase for insertion sequence element is5. Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2058(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB15497 GB:AK026530 unnamed protein product [Homo sapiens]
 Identities = 297/299 (99%), Positives = 297/299 (99%)

Query: 1   MEQILPWQNMVEVIEPFYPKAGNGRRPYPLETMLRIHCMQHWYNLSDGAMEDALYEIASM 60
           MEQILPWQNMVEVIEPFYPKAGNGRRPYPLETMLRIHCMQHWYNLSDGAMEDALYEIASM
Sbjct: 40  MEQILPWQNMVEVIEPFYPKAGNGRRPYPLETMLRIHCMQHWYNLSDGAMEDALYEIASM 99

Query: 61  RLFARLSLDSALPDRTTIMNFRHLLEQHQLARQLFKTINRWLAEAGVMMTQGTLVDATII 120
           RLFARLSLDSALPDRTTIMNFRHLLEQHQLARQLFKTINRWLAEAGVMMTQGTLVDATII
Sbjct: 100 RLFARLSLDSALPDRTTIMNFRHLLEQHQLARQLFKTINRWLAEAGVMMTQGTLVDATII 159

Query: 121 EAPSSTKNKEQQRDPEMHQTKKGNQWHFGMKAHIGVDAKSGLTHSLVTTAANEHDLNQLX 180
           EAPSSTKNKEQQRDPEMHQTKKGNQWHFGMKAHIGVDAKSGLTHSLVTTAANEHDLNQL
Sbjct: 160 EAPSSTKNKEQQRDPEMHQTKKGNQWHFGMKAHIGVDAKSGLTHSLVTTAANEHDLNQLG 219

Query: 181 NLLHGEEQFVSADAXYQGAPQREELAEVDVDWLIAERPGKVRTLKQHPRKNKTAINIEYM 240
           NLLHGEEQFVSADA YQGAPQREELAEVDVDWLIAERPGKVRTLKQHPRKNKTAINIEYM
Sbjct: 220 NLLHGEEQFVSADAGYQGAPQREELAEVDVDWLIAERPGKVRTLKQHPRKNKTAINIEYM 279

Query: 241 KASIRARVEHPFRIIKRQFGFVKARYKGLLKNDNQLAMLFTLANLFRADQMIRQWERSH 299
           KASIRARVEHPFRIIKRQFGFVKARYKGLLKNDNQLAMLFTLANLFRADQMIRQWERSH
Sbjct: 280 KASIRARVEHPFRIIKRQFGFVKARYKGLLKNDNQLAMLFTLANLFRADQMIRQWERSH 338
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2270**

A DNA sequence (GBSx2401) was identified in *S.agalactiae* <SEQ ID 7011> which encodes the amino acid sequence <SEQ ID 7012>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
```

```
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB51958 GB:AL109661 putative eukaryotic-type serine/threonine
          protein kinase [Streptomyces coelicolor A3(2)]
 Identities = 49/169 (28%), Positives = 90/169 (52%), Gaps = 6/169 (3%)

Query: 23  PTTIRVPDVSNKTVAQAKMTLENSGLKVGAIRNIESDSVSEGLVVKTDPAAGRSRREGAK 82
           P T+++PDV+   + +A+  LE+ GL+ G +    SD V+ G V+ T P +G + R G+
Sbjct: 469 PDTVKLPDVTGYKLDKARTLLEDEGLEPGMVTRAFSDEVARGFVISTKPGSGTTVRAGSA 528

Query: 83  VNLYIATPNKSFTLGNYKEHNYKDILKDLQGKGVKKSLIKVKRKINNDYTTGTILAQSLP 142
           V L + +      + +    +  +   +L+G G+K  +     ++N++Y +G + A+  P
Sbjct: 529 VAL-VVSKGSPVDVPDVTGDDLDEARAELEGAGLK--VKTADERVNSEYDSGRV-ARQTP 584

Query: 143 EGTSFNPDGNKKLTLTVAVNDPMI-MPDVTGMTVGEVIETLTDLGLDAD 190
           E      +G+  +TLTV+    MI +PDV G +V +  + L D G + D
Sbjct: 585 EPGGRAAEGD-TVTLTVSKGPRMIEVPDVVGDSVDDAKQKLEDAGFEVD 632
 Identities = 45/161 (27%), Positives = 80/161 (48%), Gaps = 4/161 (2%)

Query: 27  RVPDVSNKTVAQAKMTLENSGLKVGAIRNIESDSVSEGLVVKTDPAAGRSRREGAKVNLY 86
           +VP + +KT AQA+  L+++GL VG +R+  SD+V  G V+ TDP  G   R+   V+L
Sbjct: 405 KVPPLLSKTEAQARDRLDDAGLDVGKVRHAYSDTVERGKVISTDPGVGDRIRKNDSVSLT 464

Query: 87  IATPNKSFTLGNYKEHNYKDILKDLQGKGVKKSLIKVKRKINNDYTTGTILAQSLPEGTS 146
           ++    +  L +   +        L+ +G++  +  V R  +++   G +++     GT+
Sbjct: 465 VSDGPDTVKLPDVTGYKLDKARTLLEDEGLEPGM--VTRAFSDEVARGFVISTKPGSGTT 522

Query: 147 FNPDGNKKLTLTVAVNDPMIMPDVTGMTVGEVIETLTDLGL 187
                   + L V+   P+ +PDVTG  + E    L   GL
Sbjct: 523 VR--AGSAVALVVSKGSPVDVPDVTGDDLDEARAELEGAGL 561
```

There is also homology to SEQ ID 3026.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2271**

A DNA sequence (GBSx2402) was identified in *S.agalactiae* <SEQ ID 7013> which encodes the amino acid sequence <SEQ ID 7014>. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9311> which encodes amino acid sequence <SEQ ID 9312> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAB90561 GB:AE001058 glutamine ABC transporter, ATP-binding
          protein (glnQ) [Archaeoglobus fulgidus]
 Identities = 142/219 (64%), Positives = 178/219 (80%)

Query: 1   MDIHQGEVVVIIGPSGSGKSTFLRTMNLLEVPTKGTVTFEGIDITDKKNDIFKMREKMGM 60
           M + +GEVVVIIGPSGSGKST LR +N LE PT G +  +G+DIT+ K DI K+R+++G+
Sbjct: 24  MKVEKGEVVVIIGPSGSGKSTLLRCINRLEEPTSGKILLDGVDITNSKIDINKVRQRIGI 83
```

```
Query: 61  VFQQFNLFPNMTVLENITLSPIKTKGLSNLDAQTKAYELLEKVGLKEKANTYPASLSGGQ 120
           VFQQFNLFP++T L+N+TL+PIK K +S  +A+     LLEKVGL++KA+ YPA LSGGQ
Sbjct: 84  VFQQFNLFPHLTALQNVTLAPIKIKKMSKREAEELGMRLLEKVGLEDKADYYPAQLSGGQ 143

Query: 121 QQRIAIARGLAMNPDVLLFDEPTSALDPEMVGEVLTVMQDLAKSGMTMVIVTHEMGFARE 180
           QQR+AIAR LAMNP+V+LFDE TSALDPE+V EVL VM+ LA+ GMTMV+VTHEMGFARE
Sbjct: 144 QQRVAIARALAMNPEVMLFDEVTSALDPELVKEVLDVMKQLARDGMTMVVVTHEMGFARE 203

Query: 181 VADRVIFMDAGIIVEQGAPKEVFEQTKEIRTRDFLSKVL 219
           V DRVIFMD G+IVE+G P+++F   K  RTR FLS +L
Sbjct: 204 VGDRVIFMDGGVIVEEGKPEQIFSNPKHERTRKFLSMIL 242
```

There is also homology to SEQ ID 1186.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2272

A DNA sequence (GBSx2403) was identified in *S.agalactiae* <SEQ ID 7015> which encodes the amino acid sequence <SEQ ID 7016>. This protein is predicted to be 4-hydroxy-2-oxoglutarate aldolase (kdgA). Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1479(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14127 GB:Z99115 deoxyphosphogluconate aldolase [Bacillus subtilis]
 Identities = 21/62 (33%), Positives = 38/62 (60%), Gaps = 4/62 (6%)

Query: 3   QLMQGKIVAVIRGNSQEEAFQAAQACIKGGISAIEIAYTNSKASQVIEQLVTQYTNQEQV 62
           +L + K++AVIR   ++EA Q  ++ +  GI A+E+ YT   AS +IE      + N+E +
Sbjct: 9   RLKEAKLIAVIRSKDKQEACQQIESLLDKGIRAVEVTYTTPGASDIIE----SFRNREDI 64

Query: 63 VV 64
           ++
Sbjct: 65 LI 66
```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2273

A DNA sequence (GBSx2405) was identified in *S.agalactiae* <SEQ ID 7017> which encodes the amino acid sequence <SEQ ID 7018>. This protein is predicted to be H repeat-associated protein (rfbQRS) (b1458). Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0207(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

There is homology to SEQ ID 504.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2274**

A DNA sequence (GBSx2406) was identified in *S.agalactiae* <SEQ ID 7019> which encodes the amino acid sequence <SEQ ID 7020>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -6.74    Transmembrane     2 -  18 (   1 -  21)
   INTEGRAL    Likelihood = -3.03    Transmembrane    73 -  89 (  73 -  92)

----- Final Results -----
               bacterial membrane --- Certainty=0.3697(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

There is also homology to SEQ ID 3376.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2275**

A DNA sequence (GBSx2407) was identified in *S.agalactiae* <SEQ ID 7021> which encodes the amino acid sequence <SEQ ID 7022>. This protein is predicted to be insertion element IS1 protein InsB (insB_5). Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4280(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2276**

A DNA sequence (GBSx2409) was identified in *S.agalactiae* <SEQ ID 7023> which encodes the amino acid sequence <SEQ ID 7024>. Analysis of this protein sequence reveals the following:

```
Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3937(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2277

A DNA sequence (GBSx2410) was identified in *S.agalactiae* <SEQ ID 7025> which encodes the amino acid sequence <SEQ ID 7026>. This protein is predicted to be triosephosphate isomerase (tpi). Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.37   Transmembrane   35 -  51 (  35 -  51)

----- Final Results -----
               bacterial membrane --- Certainty=0.1150(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC43268 GB:U07640 triosephosphate isomerase [Lactococcus
          lactis]
 Identities = 50/75 (66%), Positives = 61/75 (80%)

Query: 6  IAGNWKMNKNPEEAKAFIEAVASKLPSSELVEAGIAAPALTLSTVLEAAKGSELKIAAQN 65
          IAGNWKMNK   EA+AF+EAV + LPSS+ VE+ I APAL L+ +    +GSELK+AA+N
Sbjct: 7  IAGNWKMNKTLSEAQAFVEAVKNNLPSSDNVESVIGAPALFLAPMAYLRQGSELKLAAEN 66

Query: 66 SYFENSGAFTGENSP 80
          SYFEN+GAFTGENSP
Sbjct: 67 SYFENAGAFTGENSP 81
```

There is also homology to SEQ ID 6838:.

```
Identities = 58/77 (75%), Positives = 68/77 (87%)

Query: 6  IAGNWKMNKNPEEAKAFIEAVASKLPSSELVEAGIAAPALTLSTVLEAAKGSELKIAAQN 65
          IAGNWKMNKNP+EAKAF+EAVASKLPS++LV+  +AAPA+ L T +EAAK S LK+AAQN
Sbjct: 7  IAGNWKMNKNPQEAKAFVEAVASKLPSTDLVDVAVAAPAVDLVTTIEAAKDSVLKVAAQN 66

Query: 66 SYFENSGAFTGENSPKV 82
           YFEN+GAFTGE SPKV
Sbjct: 67 CYFENTGAFTGETSPKV 83
```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2278

A DNA sequence (GBSx2412) was identified in *S.agalactiae* <SEQ ID 7027> which encodes the amino acid sequence <SEQ ID 7028>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.39   Transmembrane   96 - 112 (  96 - 112)

----- Final Results -----
               bacterial membrane --- Certainty=0.1956(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAA14368 GB:D90354 surface protein antigen precursor
            [Streptococcus sobrinus]
 Identities = 60/129 (46%), Positives = 76/129 (58%), Gaps = 18/129 (13%)

Query: 3    ISFDNSFLETVSDDSAFQADVYLQMKRIAAGQVENTYLHTVNGYVISSNTVVTHTPQPEE 62
            ++F   FL +VS DSAFQA+VYLQMKRIA G   NTY++TVNG   SSNTV T TP+P++
Sbjct: 1442 VTFKEDFLRSVSVDSAFQAEVYLQMKRIAVGTFANTYVNTVNGITYSSNTVRTSTPEPKQ 1501

Query: 63   PSPNQP--------TPPQPPIETIEPPVPASILPNTGEQES----LLGLIG--AGILLGT 108
            PSP  P         P Q       PP  A  LP TG+  +    LLGL+   AG  L
Sbjct: 1502 PSPVDPKTTTTVVFQPRQGKAYQPAPPAGAQ-LPATGDSSNAYLPLLGLVSLTAGFSL-- 1558

Query: 109  AYGLKKKEE 117
              GL++K++
Sbjct: 1559 -LGLRRKQD 1566
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2279**

A DNA sequence (GBSx2413) was identified in *S.agalactiae* <SEQ ID 7029> which encodes the amino acid sequence <SEQ ID 7030>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3691(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9359> which encodes amino acid sequence <SEQ ID 9360> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB15793 GB:Z99123 phosphotransacetylase [Bacillus subtilis]
 Identities = 131/221 (59%), Positives = 169/221 (76%), Gaps = 2/221 (0%)

Query: 6    LVDPVILGKADEVHDSLARLGFVDQDYSIIDPEQYEKFEEMKEAFVEIRKGKATMEDADR 65
            +++P+++G  +E+      L        I DP  YE  E++ +AFVE RKGKAT E A +
Sbjct: 41   VLNPIVIGNENEIQAKAKELNLTLGGVKIYDPHTYEGMEDLVQAFVERRKGKATEEQARK 100

Query: 66   LLKDVNYFGVMLVKLGLADGMVSGAIHSTADTVRPALQIIKTKPGISRTSGVFLMNRENT 125
             L D NYFG MLV  GLADG+VSGA HSTADTVRPALQIIKTK G+ +TSGVF+M R
Sbjct: 101  ALLDENYFGTMLVYKGLADGLVSGAAHSTADTVRPALQIIKTKEGVKKTSGVFIMARG-- 158

Query: 126  QERYIFADCAINIDPNAQELAEIAVNTADTAKIFDIDPKIAMLSFSTKGSAKAPQAEKVQ 185
            +E+Y+FADCAINI P++Q+LAEIA+ +A+TAK+FDI+P++AMLSFSTKGSAK+ + EKV
Sbjct: 159  EEQYVFADCAINIAPDSQDLAEIAIESANTAKMFDIEPRVAMLSFSTKGSAKSDETEKVA 218

Query: 186  EAAKIAKDLSPELAVDGELQFDAAFVPETAEIKAPNSDVAG 226
            +A KIAK+ +PEL +DGE QFDAAFVP  AE KAP+S++ G
Sbjct: 219  DAVKIAKEKAPELTLDGEFQFDAAFVPSVAEKKAPDSEIKG 259
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 7031> which encodes the amino acid sequence <SEQ ID 7032>. Analysis of this protein sequence reveals the following:

```
Possible site: 34
```

```
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3182(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 181/227 (79%), Positives = 211/227 (92%)

Query: 1   MKFEGLVDPVILGKADEVHDSLARLGFVDQDYSIIDPEQYEKFEEMKEAFVEIRKGKATM 60
           +KFEGL++P+ILG+++EV + L +LGF DQDY+II+P +Y  F++MKEAFVE+RKGKAT+
Sbjct: 38  LKFEGLLEPIILGQSEEVRNLLTKLGFADQDYTIINPNEYADFDKMKEAFVEVRKGKATL 97

Query: 61  EDADRLLKDVNYFGVMLVKLGLADGMVSGAIHSTADTVRPALQIIKTKPGISRTSGVFLM 120
           EDAD++L+DVNYFGVMLVK+GLADGMVSGAIHSTADTVRPALQIIKTKPGISRTSGVFLM
Sbjct: 98  EDADKMLRDVNYFGVMLVKMGLADGMVSGAIHSTADTVRPALQIIKTKPGISRTSGVFLM 157

Query: 121 NRENTQERYIFADCAINIDPNAQELAEIAVNTADTAKIFDIDPKIAMLSFSTKGSAKAPQ 180
           NRENT ERY+FADCAINIDP AQELAEIAVNTA+TAKIFDIDPKIAMLSFSTKGS KAPQ
Sbjct: 158 NRENTSERYVFADCAINIDPTAQELAEIAVNTAETAKIFDIDPKIAMLSFSTKGSGKAPQ 217

Query: 181 AEKVQEAAKIAKDLSPELAVDGELQFDAAFVPETAEIKAPNSDVAGK 227
            +KV+EA +IA  L+P+LA+DGELQFDAAFVPETA IKAP+S VAG+
Sbjct: 218 VDKVREATEIATGLNPDLALDGELQFDAAFVPETAAIKAPDSAVAGQ 264
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 2280**

A DNA sequence (GBSx2414) was identified in *S.agalactiae* <SEQ ID 7033> which encodes the amino acid sequence <SEQ ID 7034>. This protein is predicted to be lipopolysaccharide biosynthesis protein-related protein. Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4076(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAG19110 GB:AE005009 Vng0600c [Halobacterium sp. NRC-1]
 Identities = 57/176 (32%), Positives = 86/176 (48%), Gaps = 20/176 (11%)

Query: 1   MKVLLYLEAEEYLKKSGIGRAIKHQEKALQIAGIDYTTNPT------------------- 41
           M+ L YLEA E L+  G+  A   Q  AL+   ++    P
Sbjct: 2   MRALNYLEAAEALR-GGMVTATNQQRAALETTDVEVVETPWRAGDPVRSIGSLAAGGSCF 60

Query: 42  DDFDLVHMNTYGIRSWLLMSKAKKTGKKVIMHGHSTEEDFRNSFIGSNLVSPLFKWYLCR 101
             FD+ H N  G  S   +   A++T   +++H H T EDF  SF GS+ ++P  + YL
Sbjct: 61  TAFDVAHCNLVGPGSVAVARHARRTDTPLVLHAHLTREDFAQSFRGSSTIAPALEPYLRW 120

Query: 102 FYQKADAIITPTDYSKQLIKAYGIKKPIFVLSNGIDLSRYQXSEKKESAFRHYFHL 157
           FY +AD ++ P++Y+K +++AY +  PI  LSNG+DL   Q  E  + R  F L
Sbjct: 121 FYSQADLVLCPSEYTKDVLRAYPVDAPIRQLSNGVDLESMQGYESFRADTRARFDL 176
```

There is also homology to SEQ ID 1220.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2281**

A DNA sequence (GBSx2415) was identified in *S.agalactiae* <SEQ ID 7035> which encodes the amino acid sequence <SEQ ID 7036>. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2625(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC35010 GB:AF055987 intracellular a-amylase [Streptococcus mutans]
 Identities = 27/46 (58%), Positives = 33/46 (71%)

Query: 1   MEVGEIYAGKTFVDYLGNCEQEVVIGDDGWGDFLVESASISAWVPK 46
            M +GE    K FVDYL NC +EV++ D GWGDF V+ AS+SAWV K
Sbjct: 438 MNMGEFNRNKVFVDYLNNCTEEVILDDQGWGDFPVQEASLSAWVNK 483
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2282**

A DNA sequence (GBSx2416) was identified in *S.agalactiae* <SEQ ID 7037> which encodes the amino acid sequence <SEQ ID 7038>. This protein is predicted to be RopA. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2082(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

There is also homology to SEQ ID 6908:

```
Identities = 30/35 (85%), Positives = 33/35 (93%)

Query: 1   MEADQVRGLLSADMLKHDIAMKKAVDVITSSATVK 35
            M ADQVR LLSADMLKHDIAMKKAV+VITS+A+VK
Sbjct: 422 MPADQVRSLLSADMLKHDIAMKKAVEVITSTASVK 456
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 2283**

A DNA sequence (GBSx2417) was identified in *S.agalactiae* <SEQ ID 7039> which encodes the amino acid sequence <SEQ ID 7040>. This protein is predicted to be DNA-directed RNA polymerase, subunit delta. Analysis of this protein sequence reveals the following:

```
Possible site: 54
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2407(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB15744 GB:Z99123 RNA polymerase (delta subunit) [Bacillus subtilis]
 Identities = 62/186 (33%), Positives = 102/186 (54%), Gaps = 15/186 (8%)

Query: 1   MELEVFAGQEKSELSMIEVARAILEQRGRDNEMYFSDLVNDIQTYLGKSDSAIRESLPFF 60
           M ++ ++ +E  E++++E+A  + E+  +   + F +L+N+I + LG     + + +  F
Sbjct: 1   MGIKQYSQEELKEMALVEIAHELFEEHKKP--VPFQELLNEIASLLGVKKEELGDRIAQF 58

Query: 61  YSDLNTDGSFIPLGENKWGLRSWYAIDEIDEEIITLEEDEDGAPKRKKKRVNAFMDGDED 120
           Y+DLN DG F+ L +  WGLRSWY  D++DEE            K KKK+    ++ D D
Sbjct: 59  YTDLNIDGRFLALSDQTWGLRSWYPYDQLDEE-------TQPTVKAKKKKAKKAVEEDLD 111

Query: 121 AIDYNDDDPEDEDFTEETPSLEYDEENPDDEKSEVESYDSEINEIIPDEDLDEDVEINEE 180
             ++ + D +D D  E    L+ + ++ D+E  + +  D EI E I DED DED
Sbjct: 112 LDEFEEIDEDDLDLDEVEEELDLEADDFDEEDLDEDDDDLEIEEDIIDED-DEDY----- 165

Query: 181 DDEEEE 186
           DDEEEE
Sbjct: 166 DDEEEE 171
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 7041> which encodes the amino acid sequence <SEQ ID 7042>. Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2263(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 162/191 (84%), Positives = 181/191 (93%), Gaps = 1/191 (0%)

Query: 1   MELEVFAGQEKSELSMIEVARAILEQRGRDNEMYFSDLVNDIQTYLGKSDSAIRESLPFF 60
           ++L+VFAGQEKSELSMIEVARAILE+RGRDNEMYFSDLVN+IQ YLGKSD+ IR +LPFF
Sbjct: 12  LKLDVFAGQEKSELSMIEVARAILEERGRDNEMYFSDLVNEIQNYLGKSDAGIRHALPFF 71

Query: 61  YSDLNTDGSFIPLGENKWGLRSWYAIDEIDEEIITLEEDEDGAPKRKKKRVNAFMDGDED 120
           Y+DLNTDGSFIPLGENKWGLRSWYAIDEIDEEIITLEEDEDGA KRKKKRVNAFMDGDED
Sbjct: 72  YTDLNTDGSFIPLGENKWGLRSWYAIDEIDEEIITLEEDEDGAQKRKKKRVNAFMDGDED 131

Query: 121 AIDYNDDDPEDEDFTEETPSLEYDEENPDDEKSEVESYDSEINEIIPDEDLDEDVEINEE 180
           AIDY DDDPEDEDFTEE+  +EYDEE+PDDEKSEVESYDSE+NEIIP++D  E+V+INEE
Sbjct: 132 AIDYRDDDPEDEDFTEESAEVEYDEEDPDDEKSEVESYDSELNEIIPEDDF-EEVDINEE 190

Query: 181 DDEEEEEEEEV 191
           D+E+EE+EE V
Sbjct: 191 DEEDEEDEEPV 201
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 2284**

A DNA sequence (GBSx2418) was identified in *S.agalactiae* <SEQ ID 7043> which encodes the amino acid sequence <SEQ ID 7044>. This protein is predicted to be CTP synthetase (pyrG). Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -0.11    Transmembrane    5 -  21 (   5 -  21)

----- Final Results -----
               bacterial membrane --- Certainty=0.1044(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA09021 GB:AJ010153 CTP synthetase [Lactococcus lactis subsp.
           cremoris] (ver 2)
 Identities = 421/533 (78%), Positives = 481/533 (89%) .

Query: 2   TKYIFVTGGVVSSIGKGIVAASLGRLLKNRGLKVTIQKFDPYINIDPGTMSPYQHGEVYV 61
           TKYIFVTGG  SS+GKGIVAASLGRLLKNRGLKVT+QKFDPY+NIDPGTMSPYQHGEV+V
Sbjct: 3   TKYIFVTGGGTSSMGKGIVAASLGRLLKNRGLKVTVQKFDPYLNIDPGTMSPYQHGEVFV 62

Query: 62  TDDGAETDLDLGHYERFIDINLNKYSNVTTGKIYSEVLKKERRGEYLGATVQVIPHVTDA 121
           TDDGAETDLDLGHYERFIDINLNKYSNVT+GK+YSE+L+KER+GEYLGATVQ++PHVT+
Sbjct: 63  TDDGAETDLDLGHYERFIDINLNKYSNVTSGKVYSEILRKERKGEYLGATVQMVPHVTNM 122

Query: 122 LKEKIKRAATTTDSDVIITEVGGTVGDIESLPFLEALRQMKADVGSDNVMYIHTTLLPYL 181
           LKEKIKRAATTTD+D+IITEVGGTVGD+ESLPF+EALRQMKA+VG+DNVMYIHT  + +L
Sbjct: 123 LKEKIKRAATTTDADIIITEVGGTVGDMESLPFIEALRQMKAEVGADNVMYIHTVPILHL 182

Query: 182 KAAGEMKTKPTQHSVKELRGLGIQPNMLVIRTEQPAGQSIKNKLAQFCDVAPEAVIESLD 241
           +AAGE+KTK  Q++ K LR  GIQ NMLV+R+E P    +++K+A FCDVAPEAVI+SLD
Sbjct: 183 RAAGELKTKIAQNATKTLREYGIQANMLVLRSEVPITTEMRDKIAMFCDVAPEAVIQSLD 242

Query: 242 VDHIYQIPLNMQAQNMDQIVCDHLKLETPAADMTEWSAMVDKVMNLEKKVKIALVGKYVE 301
           V+H+YQIPLN+QAQNMDQIVCDHLKL+ P ADM EWSAMVD VMNL+KKVKIALVGKYVE
Sbjct: 243 VEHLYQIPLNLQAQNMDQIVCDHLKLDAPKADMAEWSAMVDHVMNLKKKVKIALVGKYVE 302

Query: 302 LPDAYLSVVEALKHSGYVNDVAIDLKWVNAAEVTEDNIKELVGDADGIIVPGGFGQRGSE 361
           LPDAY+SV EALKH+GY +D  +D+ WVNA +VT++N+ ELVGDA GIIVPGGFGQRG+E
Sbjct: 303 LPDAYISVTEALKHAGYASDAEVDINWVNANDVTDENVAELVGDAAGIIVPGGFGQRGTE 362

Query: 362 GKIEAIRYARENDVPMLGVCLGMQLTCVEFARNVLNLHGANSAELDPKTPFPIIDIMRDQ 421
           GKI AI+YARENDVPMLG+CLGMQLT VEFARNVL L GA+S ELDP+T +P+IDIMRDQ
Sbjct: 363 GKIAAIKYARENDVPMLGICLGMQLTAVEFARNVLGLEGAHSFELDPETKYPVIDIMRDQ 422

Query: 422 IDIEDMGGTLRLGLYPCKLKSGSRAAAAYNNQEVVQRRHRHRYEFNTKFREQFEAAGFVF 481
           +D+EDMGGTLRLGLYP KLK+GSRA AAYN+ EVVQRRHRHRYEFN K+RE FE AGFVF
Sbjct: 423 VDVEDMGGTLRLGLYPAKLKNGSRAKAAYNDAEVVQRRHRHRYEFNNKYREDFEKAGFVF 482

Query: 482 SGVSPDNRLMEVVELPEKKFFVAAQYHPELQSRPNHAEELYTAFVTAAVENMK 534
           SGVSPDNRL+E+VEL  KKFFVA QYHPELQSRPN  EELYT F+  AVEN K
Sbjct: 483 SGVSPDNRLVEIVELSGKKFFVACQYHPELQSRPNRPEELYTEFIRVAVENSK 535
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 7045> which encodes the amino acid sequence <SEQ ID 7046>. Analysis of this protein sequence reveals the following:

```
     Possible site: 23
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -0.11    Transmembrane    5 -  21 (   5 -  21)

----- Final Results -----
               bacterial membrane --- Certainty=0.1044(Affirmative) < succ>
```

```
  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAA09021 GB:AJ010153 CTP synthetase [Lactococcus lactis subsp.
          cremoris] (ver 2)
 Identities = 423/532 (79%), Positives = 483/532 (90%)

Query: 2   TKYIFVTGGVVSSIGKGIVAASLGRLLKNRGLKVTIQKFDPYINIDPGTMSPYQHGEVYV 61
           TKYIFVTGG  SS+GKGIVAASLGRLLKNRGLKVT+QKFDPY+NIDPGTMSPYQHGEV+V
Sbjct: 3   TKYIFVTGGGTSSMGKGIVAASLGRLLKNRGLKVTVQKFDPYLNIDPGTMSPYQHGEVFV 62

Query: 62  TDDGAETDLDLGHYERFIDINLNKYSNVTTGKIYSEVLRKERKGEYLGATVQVIPHITDA 121
           TDDGAETDLDLGHYERFIDINLNKYSNVT+GK+YSE+LRKERKGEYLGATVQ++PH+T+
Sbjct: 63  TDDGAETDLDLGHYERFIDINLNKYSNVTSGKVYSEILRKERKGEYLGATVQMVPHVTNM 122

Query: 122 LKEKIKRAASTTDSDVIITEVGGTVGDIESLPFLEALRQMKADVGSENVMYIHTTLLPYL 181
           LKEKIKRAA+TTD+D+IITEVGGTVGD+ESLPF+EALRQMKA+VG++NVMYIHT  + +L
Sbjct: 123 LKEKIKRAATTTDADIIITEVGGTVGDMESLPFIEALRQMKAEVGADNVMYIHTVPILHL 182

Query: 182 KAAGEMKTKPTQHSVKELRGLGIQPNMLVIRTEEPVEQGIKNKLAQFCDVNSEAVIESRD 241
           +AAGE+KTK  Q++ K LR  GIQ NMLV+R+E P+   +++K+A FCDV  EAVI+S D
Sbjct: 183 RAAGELKTKIAQNATKTLREYGIQANMLVLRSEVPITTEMRDKIAMFCDVAPEAVIQSLD 242

Query: 242 VEHLYQIPLNLQAQSMDQIVCDHLKLNAPQADMTEWSAMVDKVMNLRKTTKIALVGKYVE 301
           VEHLYQIPLNLQAQ+MDQIVCDHLKL+AP+ADM EWSAMVD VMNL+K  KIALVGKYVE
Sbjct: 243 VEHLYQIPLNLQAQNMDQIVCDHLKLDAPKADMAEWSAMVDHVMNLKKKVKIALVGKYVE 302

Query: 302 LPDAYLSVVEALKHSGYANDTAIDLKWVNANDVTVDNAADLLGDADGIIVPGGFGQRGTE 361

           LPDAY+SV EALKH+GYA+D  +D+ WVNANDVT +N A+L+GDA GIIVPGGFGQRGTE
Sbjct: 303 LPDAYISVTEALKHAGYASDAEVDINWVNANDVTDENVAELVGDAAGIIVPGGFGQRGTE 362

Query: 362 GKIQAIRYARENDVPMLGICLGMQLTCVEFARHVLNMEGANSFELEPSTKYPIIDIMRDQ 421
           GKI AI+YARENDVPMLGICLGMQLT VEFAR+VL +EGA+SFEL+P TKYP+IDIMRDQ
Sbjct: 363 GKIAAIKYARENDVPMLGICLGMQLTAVEFARNVLGLEGAHSFELDPETKYPVIDIMRDQ 422

Query: 422 IDIEDMGGTLRLGLYPCKLKPGSKAAMAYNNQEVVQRRHRHRYEFNNKFRPEFEAAGFVF 481
           +D+EDMGGTLRLGLYP KLK GS+A  AYN+ EVVQRRHRHRYEFNNK+R +FE AGFVF
Sbjct: 423 VDVEDMGGTLRLGLYPAKLKNGSRAKAAYNDAEVVQRRHRHRYEFNNKYREDFEKAGFVF 482

Query: 482 SGVSPDNRLVEIVELKEKKFFVAAQYHPELQSRPNRPEELYTAFVTAAIKNS 533
           SGVSPDNRLVEIVEL  KKFFVA QYHPELQSRPNRPEELYT F+  A++NS
Sbjct: 483 SGVSPDNRLVEIVELSGKKFFVACQYHPELQSRPNRPEELYTEFIRVAVENS 534
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 477/532 (89%), Positives = 503/532 (93%)

Query: 1   MTKYIFVTGGVVSSIGKGIVAASLGRLLKNRGLKVTIQKFDPYINIDPGTMSPYQHGEVY 60
           MTKYIFVTGGVVSSIGKGIVAASLGRLLKNRGLKVTIQKFDPYINIDPGTMSPYQHGEVY
Sbjct: 1   MTKYIFVTGGVVSSIGKGIVAASLGRLLKNRGLKVTIQKFDPYINIDPGTMSPYQHGEVY 60

Query: 61  VTDDGAETDLDLGHYERFIDINLNKYSNVTTGKIYSEVLKKERRGEYLGATVQVIPHVTD 120
           VTDDGAETDLDLGHYERFIDINLNKYSNVTTGKIYSEVL+KER+GEYLGATVQVIPH+TD
Sbjct: 61  VTDDGAETDLDLGHYERFIDINLNKYSNVTTGKIYSEVLRKERKGEYLGATVQVIPHITD 120

Query: 121 ALKEKIKRAATTTDSDVIITEVGGTVGDIESLPFLEALRQMKADVGSDNVMYIHTTLLPY 180
           ALKEKIKRAA+TTDSDVIITEVGGTVGDIESLPFLEALRQMKADVGS+NVMYIHTTLLPY
Sbjct: 121 ALKEKIKRAASTTDSDVIITEVGGTVGDIESLPFLEALRQMKADVGSENVMYIHTTLLPY 180

Query: 181 LKAAGEMKTKPTQHSVKELRGLGIQPNMLVIRTEQPAGQSIKNKLAQFCDVAPEAVIESL 240
           LKAAGEMKTKPTQHSVKELRGLGIQPNMLVIRTE+P  Q IKNKLAQFCDV  EAVIES
Sbjct: 181 LKAAGEMKTKPTQHSVKELRGLGIQPNMLVIRTEEPVEQGIKNKLAQFCDVNSEAVIESR 240

Query: 241 DVDHIYQIPLNMQAQNMDQIVCDHLKLETPAADMTEWSAMVDKVMNLEKKVKIALVGKYV 300
           DV+H+YQIPLN+QAQ+MDQIVCDHLKL  P ADMTEWSAMVDKVMNL K  KIALVGKYV
```

```
Sbjct: 241 DVEHLYQIPLNLQAQSMDQIVCDHLKLNAPQADMTEWSAMVDKVMNLRKTTKIALVGKYV 300

Query: 301 ELPDAYLSVVEALKHSGYVNDVAIDLKWVNAAEVTEDNIKELVGDADGIIVPGGFGQRGS 360
           ELPDAYLSVVEALKHSGY ND AIDLKWVNA +VT DN  +L+GDADGIIVPGGFGQRG+
Sbjct: 301 ELPDAYLSVVEALKHSGYANDTAIDLKWVNANDVTVDNAADLLGDADGIIVPGGFGQRGT 360

Query: 361 EGKIEAIRYARENDVPMLGVCLGMQLTCVEFARNVLNLHGANSAELDPKTPFPIIDIMRD 420
           EGKI+AIRYARENDVPMLG+CLGMQLTCVEFAR+VLN+ GANS EL+P T +PIIDIMRD
Sbjct: 361 EGKIQAIRYARENDVPMLGICLGMQLTCVEFARHVLNMEGANSFELEPSTKYPIIDIMRD 420

Query: 421 QIDIEDMGGTLRLGLYPCKLKSGSRAAAAYNNQEVVQRRHRHRYEFNTKFREQFEAAGFV 480
           QIDIEDMGGTLRLGLYPCKLK GS+AA AYNNQEVVQRRHRHRYEFN KFR +FEAAGFV
Sbjct: 421 QIDIEDMGGTLRLGLYPCKLKPGSKAAMAYNNQEVVQRRHRHRYEFNNKFRPEFEAAGFV 480

Query: 481 FSGVSPDNRLMEVVELPEKKFFVAAQYHPELQSRPNHAEELYTAFVTAAVEN 532
           FSGVSPDNRL+E+VEL EKKFFVAAQYHPELQSRPN  EELYTAFVTAA++N
Sbjct: 481 FSGVSPDNRLVEIVELKEKKFFVAAQYHPELQSRPNRPEELYTAFVTAAIKN 532
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 2285**

A DNA sequence (GBSx2419) was identified in *S.agalactiae* <SEQ ID 7047> which encodes the amino acid sequence <SEQ ID 7048>. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -9.92   Transmembrane   13 -  29 (   3 -  34)

----- Final Results -----
               bacterial membrane --- Certainty=0.4970(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9285> which encodes amino acid sequence <SEQ ID 9286> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14296 GB:Z99116 yqkD [Bacillus subtilis]
 Identities = 79/289 (27%), Positives = 139/289 (47%), Gaps = 8/289 (2%)

Query: 1   MKKIRLSKFIKMIVVILFLISVAASFYFFHVAQVRDDKSFISNGQRKPGNSLYAYDKSFD 60
           MKKI L+  I  +V  +  I +  S     + +  D+   I   +   G+ ++   +SF+
Sbjct: 1   MKKILLA--IGALVTAVIAIGIVFSHMILFIKKKTDED--IIKRETDNGHDVF---ESFE 53

Query: 61  KLLKQKIEMTNQNIKQVAWYVPAVKKTHKTAVVVHGFANSKENMKAYGWLFHKLGYNVLM 120
           ++ K    + +     +  Y  A   T  T ++ HG   +  N   Y  LF  LG+NVL+
Sbjct: 54  QMEKTAFVIPSAYGYDIKGYHVAPHDTPNTIIICHGVTMNVLNSLKYMHLFLDLGWNVLI 113

Query: 121 PDNIAHGESHGQLIGYGWNDRENIIKWTEMIVDK-NPSSQITLFGVSMGGATVMMASGEK 179
            D+  HG+S G+   YG+ +++++ K   ++ +K N    I + G SMG  T ++ +G
Sbjct: 114 YDHRRHGQSGGKTTSYGFYEKDDLNKVVSLLKNKTNHRGLIGIHGESMGAVTALLYAGAH 173

Query: 180 LPSQVVNIIEDCGYSSVWDELKFQAKEMYGLPAFPLLYEVSTISKIRAGFSYGQASSVEQ 239
                   I DC ++   ++L ++ +  Y LP++PLL       K+R G+   + S +
Sbjct: 174 CSDGADFYIADCPFACFDEQLAYRLRAEYRLPSWPLLPIADFFLKLRGGYRAREVSPLAV 233

Query: 240 LKKNNLPALFIHGDKDNFVPTSMVYDNYKATAGKKELYIVKGAKHAKSF 288
           + K   P LFIH   D+++P S     Y+   G K LYI +  +HA S+
Sbjct: 234 IDKIEKPVLFIHSKDDDYIPVSSTERLYEKKRGPKALYIAENGEHAMSY 282
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 7049> which encodes the amino acid sequence <SEQ ID 7050>. Analysis of this protein sequence reveals the following:

```
        Possible site: 24
   >>> Seems to have an uncleavable N-term signal seq
      INTEGRAL    Likelihood = -7.48   Transmembrane   10 -  26 (   3 -  32)

   ----- Final Results -----
                 bacterial membrane --- Certainty=0.3994(Affirmative) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB14296 GB:Z99116 yqkD [Bacillus subtilis]
 Identities = 88/295 (29%), Positives = 145/295 (48%), Gaps = 4/295 (1%)

Query: 10  LGILFLLITLISVGASFYFFHVAQIREEKSFINNKKRSTNNPLYPAEQSFDALPYEKRQL 69
            L I  L+  +I++G    F H+    ++K+  +  KR T+N  +   +SF+ +      +
Sbjct: 6   LAIGALVTAVIAIG--IVFSHMILFIKKKTDEDIIKRETDNG-HDVFESFEQMEKTAFVI 62

Query: 70  TNRGLKQVGWYLPAAQKTKKTAIVVHGFTNDKEDMKPYAMLFHDLGYNVLMPDNEAHGES 129
             +     +  Y  A   T  T I+ HG T +  +   Y  LF DLG+NVL+ D+  HG+S
Sbjct: 63  PSAYGYDIKGYHVAPHDTPNTIIICHGVTMNVLNSLKYMHLFLDLGWNVLIYDHRRHGQS 122

Query: 130 EGNLIGYGWNDRLNVMAWTDQLI-KENPESQITLFGLSMGAATVMMASGERLPAQVTSLI 188
             G    YG+ ++ ++      L  K N    I + G SMGA T ++ +G          I
Sbjct: 123 GGKTTSYGFYEKDDLNKVVSLLKNKTNHRGLIGIHGESMGAVTALLYAGAHCSDGADFYI 182

Query: 189 EDCGYASVWDELKFQAKAMYNLPAFPLLYEVSALSKIRAGFSYGEASSVKQLAKNKRPTL 248
             DC +A   ++L ++ +A Y LP++PLL       K+R G+   E S +  + K ++P L
Sbjct: 183 ADCPFACFDEQLAYRLRAEYRLPSWPLLPIADFFLKLRGGYRAREVSPLAVIDKIEKPVL 242

Query: 249 FIHGDKDDFVPTKMVYDNYKATKGPKEILIVKGAKHAKSFETNPEQYQKKIAAFL 303
            FIH   DD++P       Y+  +GPK + I +  +HA S+  N   Y+K +  FL
Sbjct: 243 FIHSKDDDYIPVSSTERLYEKKRGPKALYIAENGEHAMSYTKNRHTYRKTVQEFL 297
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 203/294 (69%), Positives = 246/294 (83%)

Query: 1   MKKIRLSKFIKMIVVILFLISVAASFYFFHVAQVRDDKSFISNGQRKPGNSLYAYDKSFD 60
            MK IR++K++ ++ +++ LISV ASFYFFHVAQ+R++KSFI+N +R   N LY  ++SFD
Sbjct: 1   MKTIRIAKYLGILFLLITLISVGASFYFFHVAQIREEKSFINNKKRSTNNPLYPAEQSFD 60

Query: 61  KLLKQKIEMTNQNIKQVAWYVPAVKKTHKTAVVVHGFANSKENMKAYGWLFHKLGYNVLM 120
             L  +K ++TN+ +KQV WY+PA +KT KTA+VVHGF N KE+MK Y  LFH LGYNVLM
Sbjct: 61  ALPYEKRQLTNRGLKQVGWYLPAAQKTKKTAIVVHGFTNDKEDMKPYAMLFHDLGYNVLM 120

Query: 121 PDNIAHGESHGQLIGYGWNDRENIIKWTEMIVDKNPSSQITLFGVSMGGATVMMASGEKL 180
            PDN AHGES G LIGYGWNDR N++ WT+ ++ +NP SQITLFG+SMG ATVMMASGE+L
Sbjct: 121 PDNEAHGESEGNLIGYGWNDRLNVMAWTDQLIKENPESQITLFGLSMGAATVMMASGERL 180

Query: 181 PSQVVNIIEDCGYSSVWDELKFQAKEMYGLPAFPLLYEVSTISKIRAGFSYGQASSVEQL 240
            P+QV ++IEDCGY+SVWDELKFQAK MY LPAFPLLYEVS +SKIRAGFSYG+ASSV+QL
Sbjct: 181 PAQVTSLIEDCGYASVWDELKFQAKAMYNLPAFPLLYEVSALSKIRAGFSYGEASSVKQL 240

Query: 241 KKNNLPALFIHGDKDNFVPTSMVYDNYKATAGKKELYIVKGAKHAKSFETEPEK 294
             KN  P LFIHGDKD+FVPT MVYDNYKAT G KE+ IVKGAKHAKSFET PE+
Sbjct: 241 AKNKRPTLFIHGDKDDFVPTKMVYDNYKATKGPKEILIVKGAKHAKSFETNPEQ 294
```

SEQ ID 9286 (GBS662) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 136 (lane 8-10; MW 63kDa) and in Figure 187 (lane 4; MW 63kDa).

GBS662-GST was purified as shown in Figure 237, lane 7.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 2286**

A DNA sequence (GBSx2420) was identified in *S.agalactiae* <SEQ ID 7051> which encodes the amino acid sequence <SEQ ID 7052>. This protein is predicted to be aspartate--ammonia ligase (asnA). Analysis of this protein sequence reveals the following:

```
Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2898(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9309> which encodes amino acid sequence <SEQ ID 9310> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC22222 GB:U32738 aspartate--ammonia ligase (asnA) [Haemophilus influenzae Rd]
 Identities = 246/300 (82%), Positives = 268/300 (89%)

Query: 1   MIDKLEIVEVQGPILSQVGDGMQDNLSGIEHPVSVKVLNIPEAEFEVVHSLAKWKRHTLA 60
           +I++L I+EVQGPILSQVG+GMQDNLSGIE  V V V  IP A FEVVHSLAKWKRHTLA
Sbjct: 23  LIEQLGIIEVQGPILSQVGNGMQDNLSGIEKAVQVNVKCIPNAVFEVVHSLAKWKRHTLA 82

Query: 61  RFGFNEGEGLFVHMKALRPDEDSLDPTHSVYVDQWDWEKVIPDGRRNLDYLKETVEKIYK 120
           RF F E EGLFVHMKALRPDEDSLDPTHSVYVDQWDWEKVIP+GRRN  YLKETV  IY+
Sbjct: 83  RFNFKEDEGLFVHMKALRPDEDSLDPTHSVYVDQWDWEKVIPEGRRNFAYLKETVNSIYR 142

Query: 121 AIRLTELAVEARFDIESILPKRITFIHTEELVEKYPDLSPKERENAIAKEYGAVFLIGIG 180
           AIRLTELAVEARFDI SILPK+ITF+H+E+LV++YPDLS KERENAI KEYGAVFLIGIG
Sbjct: 143 AIRLTELAVEARFDIPSILPKQITFVHSEDLVKRYPDLSSKERENAICKEYGAVFLIGIG 202

Query: 181 GELADGKPHDGRAPDYDDWTTPSENGFKGLNGDILVWNEQLGTAFELSSMGIRVDEDALK 240
           G+L+DGKPHDGRAPDYDDWTT SENG+KGLNGDILVWN+QLG AFELSSMGIRVDE AL+
Sbjct: 203 GKLSDGKPHDGRAPDYDDWTTESENGYKGLNGDILVWNDQLGKAFELSSMGIRVDESALR 262

Query: 241 RQVVLTGDEDRLEFEWHKTLLRGFFPLTIGGGIGQSRLAMFLLRKXHIGEVQSSVWPKEV 300
            QV LTGDED L+ +WH+ LL G  PLTIGGGIGQSRLAM LLRK HIGEVQSSVWPKE+
Sbjct: 263 LQVGLTGDEDHLKMDWHQDLLNGKLPLTIGGGIGQSRLAMLLLRKKHIGEVQSSVWPKEM 322
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 7053> which encodes the amino acid sequence <SEQ ID 7054>. Analysis of this protein sequence reveals the following:

```
     Possible site: 34
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.16   Transmembrane  189 - 205 ( 189 - 205)

----- Final Results -----
               bacterial membrane --- Certainty=0.1065(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAC22222 GB:U32738 aspartate--ammonia ligase (asnA) [Haemophilus influenzae Rd]
 Identities = 255/330 (77%), Positives = 289/330 (87%)

Query: 1   MKKSFIHQQEEISFVKNTFTQYLIAKLDVVEVQGPILSRVGDGMQDNLSGTENPVSVNVL 60
           MKK+FI QQ+EISFVKNTFTQ LI +L ++EVQGPILS+VG+GMQDNLSG E  V VNV
```

```
Sbjct: 1   MKKTFILQQQEISFVKNTFTQNLIEQLGIIEVQGPILSQVGNGMQDNLSGIEKAVQVNVK 60

Query: 61  KIPNATFEVVHSLAKWKRHTLARFGFNEGEGLVVNMKALRPDEDSLDQTHSVYVDQWDWE 120
            IPNA FEVVHSLAKWKRHTLARF F E EGL V+MKALRPDEDSLD THSVYVDQWDWE
Sbjct: 61  CIPNAVFEVVHSLAKWKRHTLARFNFKEDEGLFVHMKALRPDEDSLDPTHSVYVDQWDWE 120

Query: 121 KVIPDGKRNLAYLKETVETIYKVIRLTELAVEARYDIEAVLPKKITFIHTEELVAKYPDL 180
           KVIP+G+RN AYLKETV +IY+ IRLTELAVEAR+DI ++LPK+ITF+H+E+LV +YPDL
Sbjct: 121 KVIPEGRRNFAYLKETVNSIYRAIRLTELAVEARFDIPSILPKQITFVHSEDLVKRYPDL 180

Query: 181 TPKERENAITKEFGAVFLIGIGGVLPDGKPHDGRAPDYDDWTTETENGYHGLNGDILVWN 240
           + KERENAI KE+GAVFLIGIGG L DGKPHDGRAPDYDDWTTE+ENGY GLNGDILVWN
Sbjct: 181 SSKERENAICKEYGAVFLIGIGGKLSDGKPHDGRAPDYDDWTTESENGYKGLNGDILVWN 240

Query: 241 DQLGSAFELSSMGIRVDEEALKRQVEMTGDQDRLGFDWHKSLLNGLFPLTIGGGIGQSRM 300
           DQLG AFELSSMGIRVDE AL+ QV +TGD+D L  DWH+ LLNG  PLTIGGGIGQSR+
Sbjct: 241 DQLGKAFELSSMGIRVDESALRLQVGLTGDEDHLKMDWHQDLLNGKLPLTIGGGIGQSRL 300

Query: 301 VMFLLRKQHIGEVQTSVWPQEVRDSYDNIL 330
            M LLRK+HIGEVQ+SVWP+E+ + + NIL
Sbjct: 301 AMLLLRKKHIGEVQSSVWPKEMLEEFSNIL 330
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 254/303 (83%), Positives = 280/303 (91%)

Query: 1   MIDKLEIVEVQGPILSQVGDGMQDNLSGIEHPVSVKVLNIPEAEFEVVHSLAKWKRHTLA 60
           +I KL++VEVQGPILS+VGDGMQDNLSG E+PVSV VL IP A FEVVHSLAKWKRHTLA
Sbjct: 23  LIAKLDVVEVQGPILSRVGDGMQDNLSGTENPVSVNVLKIPNATFEVVHSLAKWKRHTLA 82

Query: 61  RFGFNEGEGLFVHMKALRPDEDSLDPTHSVYVDQWDWEKVIPDGRRNLDYLKETVEKIYK 120
           RFGFNEGEGL V+MKALRPDEDSLD THSVYVDQWDWEKVIPDG+RNL YLKETVE IYK
Sbjct: 83  RFGFNEGEGLVVNMKALRPDEDSLDQTHSVYVDQWDWEKVIPDGKRNLAYLKETVETIYK 142

Query: 121 AIRLTELAVEARFDIESILPKRITFIHTEELVEKYPDLSPKERENAIAKEYGAVFLIGIG 180
            IRLTELAVEAR+DIE++LPK+ITFIHTEELV KYPDL+PKERENAI KE+GAVFLIGIG
Sbjct: 143 VIRLTELAVEARYDIEAVLPKKITFIHTEELVAKYPDLTPKERENAITKEFGAVFLIGIG 202

Query: 181 GELADGKPHDGRAPDYDDWTTPSENGFKGLNGDILVWNEQLGTAFELSSMGIRVDEDALK 240
           G L DGKPHDGRAPDYDDWTT +ENG+ GLNGDILVWN+QLG+AFELSSMGIRVDE+ALK
Sbjct: 203 GVLPDGKPHDGRAPDYDDWTTETENGYHGLNGDILVWNDQLGSAFELSSMGIRVDEEALK 262

Query: 241 RQVVLTGDEDRLEFEWHKTLLRGFFPLTIGGGIGQSRLAMFLLRKXHIGEVQSSVWPKEV 300
           RQV +TGD+DRL F+WHK+LL G FPLTIGGGIGQSR+ MFLLRK HIGEVQ+SVWP+EV
Sbjct: 263 RQVEMTGDQDRLGFDWHKSLLNGLFPLTIGGGIGQSRMVMFLLRKQHIGEVQTSVWPQEV 322

Query: 301 RDT 303
           RD+
Sbjct: 323 RDS 325
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 2287**

A DNA sequence (GBSx2421) was identified in *S.agalactiae* <SEQ ID 7055> which encodes the amino acid sequence <SEQ ID 7056>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3163(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2288**

A DNA sequence (GBSx2422) was identified in *S.agalactiae* <SEQ ID 7057> which encodes the amino acid sequence <SEQ ID 7058>. Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9007> which encodes amino acid sequence <SEQ ID 9008> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD56628 GB:AF165218 Bta [Streptococcus pneumoniae]
 Identities = 30/97 (30%), Positives = 50/97 (50%), Gaps = 3/97 (3%)

Query: 50  KALVSKSQQSEATIFIGRPTCQYCRAFLPKLLKSQATLHSKIYYLDSQKYKG-KRLKSFF 108
           +A  +  ++  AT FIGR TC YCR F   L    A   + IY+++S++      L++F
Sbjct: 18  RAQEALDKKETATFFIGRKTCPYCRKFAGTLSGVVAETKAHIYFINSEEASQLNDLQAFR 77

Query: 109 KKHHITTVPNLAHYQQGKMTKYLVQGSQATPQQIQTF 145
            ++ I TVP   H   G++   +   S  + Q+I+ F
Sbjct: 78  SRYGIPTVPGFVHITDGQIN--VRCDSSMSAQEIKDF 112
```

SEQ ID 9008 (GBS134) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 40 (lane 2; MW 17kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 46 (lane 4; MW 42kDa).

GBS134-GST was purified as shown in Figure 204, lane 10.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2289**

A DNA sequence (GBSx2423) was identified in *S.agalactiae* <SEQ ID 7059> which encodes the amino acid sequence <SEQ ID 7060>. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0735(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9603> which encodes amino acid sequence <SEQ ID 9604> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB06309 GB:AP001516 unknown conserved protein [Bacillus halodurans]
 Identities = 78/178 (43%), Positives = 115/178 (63%), Gaps = 3/178 (1%)

Query: 3   MRVVAGTFGGRPLKTLDGKTTRPTTDKVKGAIFNMIGPFFEGGRVLDLFSGSGSLAIEAI 62
           MRV+AG   G  LK + G  TRPTTDKVK AIFNMIGPFF+GG  LDL+ GSG L IEA+
Sbjct: 1   MRVIAGEQKGLTLKAVPGHKTRPTTDKVKEAIFNMIGPFFDGGIGLDLYGGSGGLGIEAL 60

Query: 63  SRGMDQAVLVEKDRRAQVVIQENIAMTKSPEQFQLLKMEANRALEQLTGQ---FDLVLLD 119
           SRG+++ + V++ +RA   I++N++      + ++ + +A RAL+ LT +   F  V LD
Sbjct: 61  SRGVERMIFVDQQKRAIETIKQNLSHCGLEGRAEVYRNDAKRALQVLTKRGIVFAYVFLD 120

Query: 120 PPYAKEEIVKQIQIMDSKGLLGDDIMIACETDKSVDLPEEIASFGIWKQKIYGISKVT 177
           PPYAK+ I   + I+ + GLL +  ++ CE D+   LP++I      K++ YG + +T
Sbjct: 121 PPYAKQTIKNDLAILANHGLLEEGGVVVCEHDRDTMLPDQIEYAVKHKEETYGDTMIT 178
```

There is also homology to SEQ ID 132.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2290**

A DNA sequence (GBSx2424) was identified in *S.agalactiae* <SEQ ID 7061> which encodes the amino acid sequence <SEQ ID 7062>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4984(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB96619 GB:AJ400630 hypothetical protein [Streptococcus pneumoniae bacteriophage MM1]
 Identities = 175/254 (68%), Positives = 219/254 (85%)

Query: 2   LRRHIYSMLEEHXHLQPEIKYHQKTNLRKNRVYTVFIEEKVDVILADLKLADAFFGIETG 61
           L RH+Y         ++ EI++HQ++NLRKNRVYTVF +EKV  +L+DL LAD+FFG+ETG
Sbjct: 50  LARHLYESFLHFYEIKSEIRHHQRSNLRKNRVYTVFTDEKVQDLLSDLHLADSFFGLETG 109

Query: 62  IEHSILDNDENGRAYLRGAFLSTGTVREPDSGKYQLEIFSVYLDHAQDLANLMKKFMLDA 121
           I+ +IL ++E GRAYL GAFL+ G++R+P+SGKYQLEI SVYLDHAQ +A+L+++F+LDA
Sbjct: 110 IDEAILSDEEAGRAYLCGAFLANGSIRDPESGKYQLEISSVYLDHAQGIASLLQQFLLDA 169

Query: 122 KVIEHKHGAVTYLQKAEDIMDFLIVIDAMEARDAFEEIKMIRETRNDINRANNVETANIA 181
           KV+E K GAVTYLQ+AEDIMDFLIVI AM+ARD FE +K++RETRND+NRANN ETANIA
Sbjct: 170 KVLERKKGAVTYLQRAEDIMDFLIVIGAMQARDDFERVKILRETRNDLNRANNAETANIA 229

Query: 182 RTITASMKTINNIIKIMDTIGFDALPSDLRQVAQVRVAHPDYSIQQIADSLETPLSKSGV 241
           RT++ASMKTINNI KI D +G + LP DL++VAQ+R+ HPDYSIQQ+ADSL TPL+KSGV
Sbjct: 230 RTVSASMKTINNISKIKDIMGLENLPVDLQEVAQLRIQHPDYSIQQLADSLSTPLTKSGV 289

Query: 242 NHRLRKINKIADEL 255
           NHRLRKINKIADEL
Sbjct: 290 NHRLRKINKIADEL 303
```

There is also homology to SEQ ID 5540:

```
Identities = 186/254 (73%), Positives = 227/254 (89%)

Query: 2   LRRHIYSMLEEHXHLQPEIKYHQKTNLRKNRVYTVFIEEKVDVILADLKLADAFFGIETG 61
           + R+IYS++E+   + PEI+YHQKTNLRKNRVYTV++E+ V+ ILADLKLAD+FFG+ETG
```

```
Sbjct: 50  IARYIYSLIEDAYVIVPEIRYHQKTNLRKNRVYTVYVEQGVETILADLKLADSFFGLETG 109

Query: 62  IEHSILDNDENGRAYLRGAFLSTGTVREPDSGKYQLEIFSVYLDHAQDLANLMKKFMLDA 121
           IE  +L +D  GR+YL+GAFL+ G++R+P+SGKYQLEI+SVYLDHAQDLA LM+KFMLDA
Sbjct: 110 IEPQVLSDDNAGRSYLKGAFLAAGSIRDPESGKYQLEIYSVYLDHAQDLAQLMQKFMLDA 169

Query: 122 KVIEHKHGAVTYLQKAEDIMDFLIVIDAMEARDAFEEIKMIRETRNDINRANNVETANIA 181
           K IEHK GAVTYLQKAEDIMDFLI+I AM  ++ FE IK++RE RNDINRANN ETANIA
Sbjct: 170 KTIEHKSGAVTYLQKAEDIMDFLIIIGAMSCKEDFEAIKLLREARNDINRANNAETANIA 229

Query: 182 RTITASMKTINNIIKIMDTIGFDALPSDLRQVAQVRVAHPDYSIQQIADSLETPLSKSGV 241
           +TI+ASMKTINNIIKIMDTIG ++LP +L+QVAQ+RV HPDYSIQQ+AD+LE P++KSGV
Sbjct: 230 KTISASMKTINNIIKIMDTIGLESLPIELQQVAQLRVKHPDYSIQQVADALEFPITKSGV 289

Query: 242 NHRLRKINKIADEL 255
           NHRLRKINKIAD+L
Sbjct: 290 NHRLRKINKIADDL 303
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 2291**

A DNA sequence (GBSx2425) was identified in *S.agalactiae* <SEQ ID 7063> which encodes the amino acid sequence <SEQ ID 7064>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0297(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2292**

A DNA sequence (GBSx2428) was identified in *S.agalactiae* <SEQ ID 7065> which encodes the amino acid sequence <SEQ ID 7066>. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2706(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB54571 GB:AJ006393 response regulator [Streptococcus pneumoniae]
 Identities = 139/190 (73%), Positives = 166/190 (87%)

Query: 8   IKIVLVDDHEMVRLGLKSFLNLQADVEVIGEASNGLEGIKKALELRPDVVVMDLVMPEMD 67
           +KI+LVDDHEMVRLGLKS+ +LQ DVEV+GEASNG +GI  ALELRPDV+VMD+VMPEM+
Sbjct: 1   MKILLVDDHEMVRLGLKSYFDLQDDVEVVGEASNGSQGIDLALELRPDVIVMDIVMPEMN 60
```

```
Query: 68  GVEATLALLKDWPEAAILVLTSYLDNEKIYPVIEAGAKGYMLKTSSAAEILNAIRKVSRG 127
           G++ATLA+LK+WPEA IL++TSYLDNEKI PV++AGAKGYMLKTSSA E+L+A+ KV+ G
Sbjct: 61  GIDATLAILKEWPEAKILIVTSYLDNEKIMPVLDAGAKGYMLKTSSADELLHAVSKVAAG 120

Query: 128 EQAIENEVDKKIKAHDKCPALHEGLTARERDILNLLAKGYDNQRIADELFISLKTVKTHV 187
           E AIE EV KK++ H     LHE LTARERD+L L+AKGY+NQRIAD+LFISLKTVKTHV
Sbjct: 121 ELAIEQEVSKKVEYHRNHMELHEELTARERDVLQLIAKGYENQRIADDLFISLKTVKTHV 180

Query: 188 SNILGKLNGS 197
           SNIL KL  S
Sbjct: 181 SNILAKLEVS 190
```

There is also high homology to SEQ ID 2996:

```
Identities = 158/198 (79%), Positives = 176/198 (88%), Gaps = 1/198 (0%)

Query: 5   MDKIKIVLVDDHEMVRLGLKSFLNLQADVEVIGEASNGLEGIKKALELRPDVVVMDLVMP 64
           M KIK++LVDDHEMVR+GLKSFLNLQAD++V+GEASNG EG+  AL L+PDV+VMDLVMP
Sbjct: 3   MSKIKVILVDDHEMVRMGLKSFLNLQADIDVVGEASNGREGVDLALALKPDVLVMDLVMP 62

Query: 65  EMDGVEATLALLKDWPEAAILVLTSYLDNEKIYPVIEAGAKGYMLKTSSAAEILNAIRKV 124
           E+ GVEATL +LK W EA +LVLTSYLDNEKIYPVI+AGAKGYMLKTSSAAEILNAIRKV
Sbjct: 63  ELGGVEATLEVLKKWKEAKVLVLTSYLDNEKIYPVIDAGAKGYMLKTSSAAEILNAIRKV 122

Query: 125 SRGEQAIENEVDKKIKAHDKCPALHEGLTARERDILNLLAKGYDNQRIADELFISLKTVK 184
           S+GE AIE EVDKKIKAHD+ P LHE LTARE DIL+LLAKGYDNQ IADELFISLKTVK
Sbjct: 123 SKGELAIETEVDKKIKAHDQHPDLHEELTAREYDILHLLAKGYDNQTIADELFISLKTVK 182

Query: 185 THVSNILGKLN-GSRSNS 201
           THVSNIL KL  G R+ +
Sbjct: 183 THVSNILAKLEVGDRTQA 200
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 2293**

A DNA sequence (GBSx2429) was identified in *S.agalactiae* <SEQ ID 7067> which encodes the amino acid sequence <SEQ ID 7068>. This protein is predicted to be histidine kinase (narQ). Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3944(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB54570 GB:AJ006393 histidine kinase [Streptococcus pneumoniae]
 Identities = 32/55 (58%), Positives = 49/55 (88%)

Query: 1   MIDNGIGFDMDSVYDLSYGLKNIEDRVEDLAGNLQLLSQPGKGVAMDIRLPLVNQ 55
           ++DNGIGF + S+ DLSYGL+NI++RVED+AG +QLL+ P +G+A+DIR+PL+++
Sbjct: 276 VVDNGIGFQLGSLDDLSYGLRNIKERVEDMAGTVQLLTAPKQGLAVDIRIPLLDK 330
```

There is also homology to SEQ ID 2992:

```
Identities = 44/59 (74%), Positives = 51/59 (85%)

Query: 1   MIDNGIGFDMDSVYDLSYGLKNIEDRVEDLAGNLQLLSQPGKGVAMDIRLPLVNQSEDK 59
           MID+G+GFDMD V DLSYGLKNIEDRV DLAGNL L+SQ GKGV+MDIRLP+V   +D+
Sbjct: 276 MIDDGVGFDMDQVRDLSYGLKNIEDRVNDLAGNLHLISQKGKGVSMDIRLPIVKGDDDE 334
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 2294**

A DNA sequence (GBSx2430) was identified in *S.agalactiae* <SEQ ID 7069> which encodes the amino acid sequence <SEQ ID 7070>. This protein is predicted to be RfbQRSO155-1. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1120(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

There is also homology to SEQ ID 7072:

```
Identities = 171/172 (99%), Positives = 172/172 (99%)

Query: 1   MGQVAVEEKSNEIVAIPQLLRTIDIRKSIVTIDAMGTQTAIVDTIIKGKADYCLAVKGNQ 60
            +GQVAVEEKSNEIVAIPQLLRTIDIRKSIVTIDAMGTQTAIVDTIIKGKADYCLAVKGNQ
Sbjct: 143 LGQVAVEEKSNEIVAIPQLLRTIDIRKSIVTIDAMGTQTAIVDTIIKGKADYCLAVKGNQ 202

Query: 61  ETLYDDIALYFSDVNLLEELQENAQYYQTVEKSRGQIEVREYWVSSDIKWLCQNHPKWHK 120
            ETLYDDIALYFSDVNLLEELQENAQYYQTVEKSRGQIEVREYWVSSDIKWLCQNHPKWHK
Sbjct: 203 ETLYDDIALYFSDVNLLEELQENAQYYQTVEKSRGQIEVREYWVSSDIKWLCQNHPKWHK 262

Query: 121 LRGIGMTRNTIDKDGQLSQENRYFIFSFKPDVLTFANCVRGHWQIESMHWLL 172
            LRGIGMTRNTIDKDGQLSQENRYFIFSFKPDVLTFANCVRGHWQIESMHWLL
Sbjct: 263 LRGIGMTRNTIDKDGQLSQENRYFIFSFKPDVLTFANCVRGHWQIESMHWLL 314
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 2295**

A DNA sequence (GBSx2431) was identified in *S.agalactiae* <SEQ ID 7073> which encodes the amino acid sequence <SEQ ID 7074>. This protein is predicted to be translation initiation factor if-3 homolog dsg (infC). Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1787(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA68920 GB:Y07640 translation initiation factor, IF3 [Listeria monocytogenes]
 Identities = 112/169 (66%), Positives = 134/169 (79%)

Query: 7   KDLFINDEIRVREVRLVGLEGEQLGIKPLSEAQAIADDANVDLVLIQPQATPPVAKIMDY 66
            KD+ +ND IR REVRL+  +GEQLG+K   +A  IA+ AN+DLVL+ P A PPVA+IMDY
Sbjct: 3   KDMLVNDGIRAREVRLIDQDGEQLGVKSKIDALQIAEKANLDLVLVAPTAKPPVARIMDY 62

Query: 67  GKFKFEYQKKQKEQRKKQSVVTVKEVRLSPVIDKGDFETKLRNGRKFLEKGNKVKVSIRF 126
            GKF+FE QKK KE RK Q V+ +KEVRLSP ID+ DF+TKLRN RKFLEKG+KVK SIRF
```

```
Sbjct: 63  GKFRFEQQKKDKEARKNQKVIVMKEVRLSPTIDEHDFDTKLRNARKFLEKGDKVKCSIRF 122

Query: 127 KGRMITHKEIGAKVLAEFAEATQDIAIIEQRAKMDGRQMFMQLAPIPDK 175
           KGR ITHKEIG KVL  FA+A +D+  IEQR KMDGR MF+ LAP+ +K
Sbjct: 123 KGRAITHKEIGQKVLDRFAKACEDLCTIEQRPKMDGRSMFLVLAPLHEK 171
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 7075> which encodes the amino acid sequence <SEQ ID 7076>. Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2247(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 167/176 (94%), Positives = 173/176 (97%)

Query: 1   MKIIAKKDLFINDEIRVREVRLVGLEGEQLGIKPLSEAQAIADDANVDLVLIQPQATPPV 60
           +KIIAKKDLFINDEIRVREVRLVGLEGEQLGIKPLSEAQ++AD +NVDLVLIQPQA PPV
Sbjct: 1   VKIIAKKDLFINDEIRVREVRLVGLEGEQLGIKPLSEAQSLADASNVDLVLIQPQAVPPV 60

Query: 61  AKIMDYGKFKFEYQKKQKEQRKKQSVVTVKEVRLSPVIDKGDFETKLRNGRKFLEKGNKV 120
           AK+MDYGKFKFEYQKKQKEQRKKQSVVTVKEVRLSPVIDKGDFETKLRNGRKFLEKGNKV
Sbjct: 61  AKLMDYGKFKFEYQKKQKEQRKKQSVVTVKEVRLSPVIDKGDFETKLRNGRKFLEKGNKV 120

Query: 121 KVSIRFKGRMITHKEIGAKVLAEFAEATQDIAIIEQRAKMDGRQMFMQLAPIPDKK 176
           KVSIRFKGRMITHKEIGAKVLA+FAEATQDIAIIEQRAKMDGRQMFMQLAPI DKK
Sbjct: 121 KVSIRFKGRMITHKEIGAKVLADFAEATQDIAIIEQRAKMDGRQMFMQLAPISDKK 176
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 2296**

A DNA sequence (GBSx2432) was identified in *S.agalactiae* <SEQ ID 7077> which encodes the amino acid sequence <SEQ ID 7078>. Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1807(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC45308 GB:U81957 RNA polymerase beta' subunit [Streptococcus gordonii]
 Identities = 262/286 (91%), Positives = 276/286 (95%)

Query: 1   MAAKVVKAGVEEVXIRSVFTCNTRHGVCRHCYGINLATGDAVEVGEAVGTIAAQSIGEPG 60
           MA +VV AGV EV IRSV TCNTRHGVCRHCYGINLATGDAVEVGEAVGTIAAQSIGEPG
Sbjct: 122 MARQVVNAGVTEVTIRSVLTCNTRHGVCRHCYGINLATGDAVEVGEAVGTIAAQSIGEPG 181

Query: 61  TQLTMRTFHTGGVASNTDITQGLPRIQEIFEARNPKGEAVITEVKGEVVAIEEDSSTRTK 120
           TQLTMRTFHTGGVAS++DITQGLPR+QEIFEARNPKGEAVITEVKGEV AIEED+STRTK
Sbjct: 182 TQLTMRTFHTGGVASSSDITQGLPRVQEIFEARNPKGEAVITEVKGEVTAIEEDASTRTK 241

Query: 121 KVFVKGQTGEGEYVVPFTARMKVEVGDEVARGAALTEGSIQPKRLLEVRDTLSVETYLLA 180
           KVFVKGQTGEGEYVVPFTARMKVEVGD+V+RGAALTEGSIQPK LL VRD LSVETYLLA
Sbjct: 242 KVFVKGQTGEGEYVVPFTARMKVEVGDQVSRGAALTEGSIQPKHLLAVRDVLSVETYLLA 301
```

```
Query: 181 EVQKVYRSQGVEIGDKHVEVMVRQMLRKVRVMDPGDTDLLPGTLMDISDFTDANKDIVIS 240
           EVQKVYRSQGVEIGDKH+EVMVRQM+RKVRVMDPGDTDLL GTLMDI+DFTDAN+D+VIS
Sbjct: 302 EVQKVYRSQGVEIGDKHIEVMVRQMIRKVRVMDPGDTDLLMGTLMDITDFTDANRDVVIS 361

Query: 241 GGIPATSRPVLMGITKASLETNSFLSAASFQETTRVLTDAAIRGKK 286
           GG+PAT+RPVLMGITKASLETNSFLSAASFQETTRVLTDAAIRGKK
Sbjct: 362 GGVPATARPVLMGITKASLETNSFLSAASFQETTRVLTDAAIRGKK 407
```

There is also homology to SEQ ID 384.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2297**

A DNA sequence (GBSx2434) was identified in *S.agalactiae* <SEQ ID 7079> which encodes the amino acid sequence <SEQ ID 7080>. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0352(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2298**

A DNA sequence (GBSx2435) was identified in *S.agalactiae* <SEQ ID 7081> which encodes the amino acid sequence <SEQ ID 7082>. This protein is predicted to be acetoin dehydrogenase (TPP-dependent) beta chain (pdhB). Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0266(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB04496 GB:AP001509 acetoin dehydrogenase (TPP-dependent) beta
           chain [Bacillus halodurans]
 Identities = 37/57 (64%), Positives = 50/57 (86%)

Query: 1   MLEEFGAKRVRDTPISEAAIAGSAIGAAQTGLRPIVDLTFMDFVTIAMDAIVDDCIR 57
           M+EEFG++RVR+TPISEAAI+G+AIGAA TG+RPI++L F DF+TIAMD +V+   +
Sbjct: 44  MIEEFGSERVRNTPISEAAISGTAIGAALTGMRPILELQFSDFITIAMDNMVNQAAK 100
```

There is also homology to SEQ ID 4272.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2299**

A DNA sequence (GBSx2436) was identified in *S.agalactiae* <SEQ ID 7083> which encodes the amino acid sequence <SEQ ID 7084>. This protein is predicted to be Structural protein. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3015(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAB18706 GB:U38906 Structural protein [Bacteriophage r1t]
 Identities = 57/127 (44%), Positives = 83/127 (64%)

Query: 5   IKAGTLFKPELVTEIMSKVKGHSTLAKLSGQTPIPFNGVEQFVFNLDGNAQIVGEGEQKL 64
           +  GTLF P LVT+++SKV G S++A+LS Q PIPFNG + F F +D    +V E  +K
Sbjct: 3   LNKGTLFDPTLVTDLISKVAGKSSIARLSAQKPIPFNGEKVFTFTMDSEIDVVAESGKKT 62

Query: 65  GNTAKVTSKIIKPLKFVYQARMTDEFKYASEEKRLNFLKHYADGFAKKMAEAFDIAAIHG 124
                +  + + P+K  Y AR++DEF YAS+E+++N L+ + DGFAKK+A   D+ A HG
Sbjct: 63  HGGVTLAPQTMVPIKVEYGARISDEFMYASDEEKINILQEFNDGFAKKVARGIDLMAFHG 122

Query: 125 LEPRTMT 131
           + PR  T
Sbjct: 123 VNPRLGT 129
```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2300**

A DNA sequence (GBSx2439) was identified in *S.agalactiae* <SEQ ID 7085> which encodes the amino acid sequence <SEQ ID 7086>. This protein is predicted to be surface protein Rib. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1892(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2301**

A DNA sequence (GBSx2440) was identified in *S.agalactiae* <SEQ ID 7087> which encodes the amino acid sequence <SEQ ID 7088>. Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.2227(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2302**

A DNA sequence (GBSx2441) was identified in *S.agalactiae* <SEQ ID 7089> which encodes the amino acid sequence <SEQ ID 7090>. This protein is predicted to be integrase. Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.2948(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9319> which encodes amino acid sequence <SEQ ID 9320> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB96616 GB:AJ400629 integrase [Streptococcus pneumoniae
           bacteriophage MM1]
 Identities = 84/238 (35%), Positives = 137/238 (57%), Gaps = 8/238 (3%)

Query: 1   MTLDKNSSQAQKKAGLILQEKIEDRLAIRNHSEMTYGELKKEYLKQWIPTVKDSTKRGYL 60
           +T++K + QA+ +A ++LQEKI  +L+ +    +T+ E+   + K W  TVK+STK
Sbjct: 30  VTMEKKTPQARNQAAILLQEKINKKLSTKQVESITFEEIYNLFYKSWAQTVKESTKHNCK 89

Query: 61  VSDSHIATVLPDDTIINKLTKRDIRLIIDKLLKHNSYHVTHKCRKRLHAIFSYAIQMDYM 120
             D  +  V+P DTI+  L +R ++  I+K+++ N Y    K R RL  IF+YA+Q  Y+
Sbjct: 90  SVDKKMKEVIPSDTILANLDRRFLQEAIEKIIESNGYITAKKVRHRLRGIFNYAVQYSYI 149

Query: 121 TSNPTENVLVP-KPK--DDYKPEKVLYLTSNEV---YDLCNRMIDNDEQTLADIVLFMFL 174
            +N  +   +P KPK  ++ + ++  +LT  E+    D+ NR      Q  AD+VL + L
Sbjct: 150 ENNEVDYTTIPQKPKTLEELEKKRNNFLTMQEIKALVDVLNRR--EYHQKYADMVLVLTL 207

Query: 175 TGVRYGELSCLTYDKIDFENKEILINATYDFNTRXITTTKTKKSTRKISVSDNILDIV 232
           TG+RYGEL+ L    IDFEN +I I   +D   + T  KT  S R I VS+++++ +
Sbjct: 208 TGMRYGELTALQLKNIDFENNKIEITGNFDSVNKIKTLPKTTNSIRTIKVSESVIEAI 265
```

There is also homology to SEQ ID 578.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2303**

A DNA sequence (GBSx2444) was identified in *S.agalactiae* <SEQ ID 7091> which encodes the amino acid sequence <SEQ ID 7092>. Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2518(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

There is also homology to SEQ ID 4212:

```
Identities = 92/144 (63%), Positives = 118/144 (81%), Gaps = 1/144 (0%)

Query: 1    MPKYSLFELENGRRRLLASAGELQKGNELALPTQFMKFLYLASRYNESKGKPEEIEKKQE 60
            +PKYSLFELENGR+R+LASAGELQKGNELALP++++ FLYLAS Y + KG PE+ E+KQ
Sbjct: 1198 LPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL 1257

Query: 61   FVNQHVSYFDDILQLINDFSKRVILADANLEKINKLYQDNKENISVDELANNIINLFTFT 120
            FV QH  Y D+I++ I++FSKRVILADANL+K+   Y   +++   + E A NII+LFT T
Sbjct: 1258 FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDK-PIREQAENIIHLFTLT 1316

Query: 121  SLGAPAAFKFFDKIVDRKRYTSTQ 144
            +LGAPAAFK+FD  +DRKRYTST+
Sbjct: 1317 NLGAPAAFKYFDTTIDRKRYTSTK 1340
```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

## Example 2304

A DNA sequence (GBSx2445) was identified in *S.agalactiae* <SEQ ID 7093> which encodes the amino acid sequence <SEQ ID 7094>. This protein is predicted to be (). Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -4.57    Transmembrane  239 - 255 ( 236 - 256)

----- Final Results -----
               bacterial membrane --- Certainty=0.2826(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB15253 GB:Z99120 similar to opine catabolism [Bacillus subtilis]
 Identities = 88/257 (34%), Positives = 129/257 (49%), Gaps = 11/257 (4%)

Query: 1    MARLGADFYSKLVTDLQKDGFETKFYQQTGVFLLKKDESQLESLFALADKRRLESPLIGD 60
            +A+ GA +Y  L+  L+KDG     Y++ G   +  D S+L+ +   A KRR ++P IGD
Sbjct: 61   LAKGGARYYKDLIHQLEKDGESDTGYKRVGAISIHTDASKLDKMEERAYKRREDAPEIGD 120

Query: 61   LQILNKSEANTHFPEL-DGYEQLLYASGGARVEGADLTRILLEAS---GVNVIKDEVHF- 115
            +  L+ SE    FP L DGYE  ++ SG ARV G  L R LL A+   G  VIK
Sbjct: 121  ITRLSASETKKLFPILADGYES-VHISGAARVNGRALCRSLLSAAEKRGATVIKGNASLL 179

Query: 116  ----TITDNGFRVQGIDFDKLVLASGAWLAKILDEHNYQVDVRPQKGQLRDYYFSNINTG 171
                T+T      +    D +++ +GAW  +IL        V  QK Q+  +  ++ +TG
Sbjct: 180  FENGTVTGVQTDTKQFAADAVIVTAGAWANEILKPLGIHFQVSFQKAQIMHFEMTDADTG 239

Query: 172  KYPVVMPEGELDIIPFDNGKVSVGASHENDMAF-DLNIDFKVLDKFEEQAIGYFPQLKKQ 230
             +PVVMP  +  I+ FDNG++  GA+HEND    DL +      +   +A+   P L
Sbjct: 240  SWPVVMPPSDQYILSFDNGRIVAGATHENDAGLDDLRVTAGGQHEVLSKALAVAPGLADA 299

Query: 231  IRLLKRVEFVPIQVIFL 247
              +  RV F P    FL
```

```
Sbjct: 300 AAVETRVGFRPFTPGFL 316
```

There is also homology to SEQ ID 2656.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2305**

A DNA sequence (GBSx2446) was identified in *S.agalactiae* <SEQ ID 7095> which encodes the amino acid sequence <SEQ ID 7096>. Analysis of this protein sequence reveals the following:

```
Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2572(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9315> which encodes amino acid sequence <SEQ ID 9316> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC00337 GB:AF008220 YtqI [Bacillus subtilis]
 Identities = 119/256 (46%), Positives = 174/256 (67%), Gaps = 3/256 (1%)

Query: 6   QILDKIKEYDTIIIHRHMRPDPDALGSQIGLRDIIRHNFPKKKVLATGFDEPTLAWIAKM 65
           +++  I  YDTII+HRH+RPDPDA GSQ GL +I+R  +P+K + A G  EP+L+++  +
Sbjct: 4   ELIRTISLYDTIILHRHVRPDPDAYGSQCGLTEILRETYPEKNIFAVGTPEPSLSFLYSL 63

Query: 66  DQVTDQDYQGALVVVTDTANTPRIDDERYKKGDFLIKIDHHPNDEVYGDLSYVDTNASSA 125
           D+V ++ Y+GALV+V DTAN  RIDD+RY  G  L+KIDHHPN++ YGDL +VDT+ASS
Sbjct: 64  DEVDNETYEGALVIVCDTANQERIDDQRYPSGAKLMKIDHHPNEDPYGDLLWVDTSASSV 123

Query: 126 SEIVTDFAL---SCDLLLSTSAARVLYNGIVGDTGRFLYPATTSKTLKIASKLREFDFDF 182
           SE++ +  L        L+T AA ++Y GIVGDTGRFL+P TT KTLK A +L ++ F
Sbjct: 124 SEMIYELYLEGKEHGWKLNTKAAELIYAGIVGDTGRFLFPNTTEKTLKYAGELIQYPFSS 183

Query: 183 SAMARQMDSFPFKIAKLQGFIFEQLKIDKNGAACVTLTQEDLKRFDVTDAETAAIVGVPG 242
           S +  Q+      + KL GFIF+ + + +NGAA V + ++ L++F  T +E + +VG  G
Sbjct: 184 SELFNQLYETKLNVVKLNGFIFQNVSLSENGAASVFIKKDTLEKFGTTASEASQLVGTLG 243

Query: 243 KIDIVESWAIFVKQSD 258
            I  + +W  FV++ D
Sbjct: 244 NISGIRAWVFFVEEDD 259
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 7097> which encodes the amino acid sequence <SEQ ID 7098>. Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2584(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 180/256 (70%), Positives = 215/256 (83%)

Query: 4   FQQILDKIKEYDTIIIHRHMRPDPDALGSQIGLRDIIRHNFPKKKVLATGFDEPTLAWIA 63
```

```
           F+ ILDKIK + TIIIHRH  PDPDALGSQ GL++II  NFP KKVL TGFDEP+LAWI+
Sbjct: 5   FETILDKIKAHQTIIIHRHQNPDPDALGSQAGLKEIIAQNFPDKKVLMTGFDEPSLAWIS 64

Query: 64  KMDQVTDQDYQGALVVVTDTANTPRIDDERYKKGDFLIKIDHHPNDEVYGDLSYVDTNAS 123
           +MDQVTD+DY+ ALV++TDTAN PRIDDERY  G  LIKIDHHPND+VYGD  YVDT+AS
Sbjct: 65  QMDQVTDKDYKEALVIITDTANRPRIDDERYTLGKCLIKIDHHPNDDVYGDFYYVDTSAS 124

Query: 124 SASEIVTDFALSCDLLLSTSAARVLYNGIVGDTGRFLYPATTSKTLKIASKLREFDFDFS 183
           SASEI+ DFA S +L LS  AA++LY GIVGDTGRFLY +TTSKTL IAS+LR F+FDF+
Sbjct: 125 SASEIIADFAFSQNLTLSDKAAKLLYTGIVGDTGRFLYASTTSKTLSIASQLRHFEFDFA 184

Query: 184 AMARQMDSFPFKIAKLQGFIFEQLKIDKNGAACVTLTQEDLKRFDVTDAETAAIVGVPGK 243
           A++RQMDSFP KIAKLQ ++FE L ID++GAA V ++QE LK FDVT AE++AIV  PGK
Sbjct: 185 AISRQMDSFPLKIAKLQSYVFEHLTIDESGAAYVLVSQETLKHFDVTLAESSAIVCAPGK 244

Query: 244 IDIVESWAIFVKQSDG 259
           ID V++WAIFV+ +DG
Sbjct: 245 IDNVQAWAIFVELTDG 260
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 2306**

A DNA sequence (GBSx2447) was identified in *S.agalactiae* <SEQ ID 7099> which encodes the amino acid sequence <SEQ ID 7100>. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1846(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB42949 GB:AL049863 putative adenosine deaminase [Streptomyces
           coelicolor A3(2)]
 Identities = 123/343 (35%), Positives = 175/343 (50%), Gaps = 26/343 (7%)

Query: 6   LKELAKAELHCHLDGSLSLPAIRKLANMADIILPSSDK-ELRKYVIAPAQTESLVDYLKT 64
           L+ L KA LH HLDG L    + +LA      LP++D  EL  +    A +  LV Y+ T
Sbjct: 11  LRRLPKAVLHDHLDGGLRPATVVELARSVGHTLPTTDPDELAAWYYEAANSGDLVRYIAT 70

Query: 65  FEFIRPLLQTKEALRFAAYDVARQAALENVIYIEIRFAPELSMDKGLTASDTVLAVLEGL 124
           FE    ++Q +E L  AA +     A + V+Y E+R+APEL+   GL+  + V  V EGL
Sbjct: 71  FEHTLAVMQNREGLLRAAEEYVLDLAADGVVYGEVRYAPELNTRGGLSMREVVETVQEGL 130

Query: 125 ADAQKEFNIVAR-----ALVCGMRQSSHKTTKDIIKHIVDLA----PKGLVGFDFAGDEF 175
           A    +            L+CGMR        D ++   DLA      G+VGFD AG E
Sbjct: 131 ATGMAKAAAAGTPVRVGTLLCGMRMF------DRVREAADLAVAFRDAGVVGFDIAGAED 184

Query: 176 SYPTDSLVDLIQEVKRSGYPMTLHAGECGCAKHIADSLNL-GIKRMGHVTALT------- 227
            +P    +D  + ++R   P T+HAGE      I  +L + G +R+GH   +T
Sbjct: 185 GFPPADHLDAFEHLRRENVPFTIHAGEAHGLPSIHQALQVCGAQRIGHGVRITDDIPDLA 244

Query: 228 -GQRDLIKRFVEEDAVA-EMCLTSNLQTKAASSIQSFPYQELYDAGGKITINTDNRTVSD 285
            G+   +  +V +  +A EMC TSNLQT AA+SI   P   L D G ++T+NTDNR VS
Sbjct: 245 AGKLGRLAAWVRDRRIALEMCPTSNLQTGAATSIAEHPITALKDLGFRVTLNTDNRLVSG 304

Query: 286 TNLTKEYSLFVTYFGTKIEDFLVFNQNAVKASFTSDSEKDTLL 328
           T +T+E SL V   G  +ED      NA+K++F    E+  L+
Sbjct: 305 TTMTREMSLLVEQAGWSVEDLRTVTVNALKSAFVPFDERTALI 347
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2307**

A DNA sequence (GBSx2448) was identified in *S.agalactiae* <SEQ ID 7101> which encodes the amino acid sequence <SEQ ID 7102>. Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2042(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9639> which encodes amino acid sequence <SEQ ID 9640> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB13290 GB:Z99111 similar to sulfite reductase [Bacillus subtilis]
 Identities = 63/146 (43%), Positives = 87/146 (59%), Gaps = 1/146 (0%)

Query: 5    MALAKIVYASMTGNTEEIADIVADKLRDLGLDVEVEECTMVDAAD-FEDADIAIVATYTY 63
            MA   +VYA+M+GNTE +AD++   L++   +V+  E   +D A  F D D  I+ TYT+
Sbjct: 1    MAKILLVYATMSGNTEAMADLIEKGLQEALAEVDRFEAMDIDDAQLFTDYDHVIMGTYTW 60

Query: 64   GDGDLPDEIVDFYEDLAEVDLSGKVYGVVGSGDTFYDYFCKSVDEFEAQFALTGAQKGAD 123
            GDGDLPDE +D  ED+ E+D SGK   V GSGDT Y++FC +VD  EA+    G
Sbjct: 61   GDGDLPDEFLDLVEDMEEIDFSGKTCAVFGSGDTAYEFFCGAVDTLEAKIKERGGDIVLP 120

Query: 124 CVKVDLAAEDEDIENLEAFAEEIASK 149
             VK++   E E+ E L  F  + A K
Sbjct: 121 SVKIENNPEGEEEEELINFGRQFAKK 146
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 7103> which encodes the amino acid sequence <SEQ ID 7104>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1641(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 116/147 (78%), Positives = 136/147 (91%)

Query: 5    MALAKIVYASMTGNTEEIADIVADKLRDLGLDVEVEECTMVDAADFEDADIAIVATYTYG 64
            MALAKIVYASMTGNTEEIADIVA+KL++LG DV+++ECT VDA++FE+ADIA+VATYTYG
Sbjct: 1    MALAKIVYASMTGNTEEIADIVANKLQELGHDVDIDECTTVDASEFENADIAVVATYTYG 60

Query: 65   DGDLPDEIVDFYEDLAEVDLSGKVYGVVGSGDTFYDYFCKSVDEFEAQFALTGAQKGADC 124
            DGDLPDEIVDFYEDL ++DL GK+YGVVGSGDTFYDYFCKSVD+F  QFALTGA KGA+
Sbjct: 61   DGDLPDEIVDFYEDLQDLDLEGKIYGVVGSGDTFYDYFCKSVDDFSEQFALTGAIKGAEP 120

Query: 125 VKVDLAAEDEDIENLEAFAEEIASKLN 151
            VKVDLAAEDEDI+ LEAFAE+++  +N
Sbjct: 121 VKVDLAAEDEDIDRLEAFAEQLSQAVN 147
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 2308**

A DNA sequence (GBSx2449) was identified in *S.agalactiae* <SEQ ID 7105> which encodes the amino acid sequence <SEQ ID 7106>. Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3568(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAB98234 GB:U67480 chorismate mutase/prephenate dehydratase
          (pheA) [Methanococcus jannaschii]
 Identities = 26/85 (30%), Positives = 46/85 (53%), Gaps = 1/85 (1%)

Query: 2   ELEEIRQEIDEIDQQLVSLLETRMGLILEVIAFKKKHRLPVLDNNRENEVLNNVLKKVQN 61
           +L EIR++IDEID +++ L+  R  L  +V   K +  +P+ D  RE  + + + K  +
Sbjct: 4   KLAEIRKKIDEIDNKILKLIAERNSLAKDVAEIKNQLGIPINDPEREKYIYDRIRKLCKE 63

Query: 62  HQFDDVIRATFKDIMTE-SRVYQKE 85
           H  D+ I      I+ E ++  QK+
Sbjct: 64  HNVDENIGIKIFQILIEHNKALQKQ 88
```

There is also homology to SEQ ID 1568.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2309**

A DNA sequence (GBSx2450) was identified in *S.agalactiae* <SEQ ID 7107> which encodes the amino acid sequence <SEQ ID 7108>. This protein is predicted to be a minor structural protein. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1828(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC34413 GB:AF158600 putative minor structural protein
           [Streptococcus thermophilus bacteriophage Sfi11]
 Identities = 39/65 (60%), Positives = 54/65 (83%)

Query: 1    MEVETDSQEVLMSTGLKDLKAHAYPAITYEVDGYVDLELGDVVRIQDDGYEPPLILTARV 60
            ME++TDS++VL+ST L++L+   YPAITYEVDG++DL++GD V+IQD G+ P L+L ARV
Sbjct: 707 MEIDTDSEDVLISTALRNLRKFCYPAITYEVDGFLDLDIGDTVKIQDTGFSPMLMLEARV 766

Query: 61   VEQDI 65
             EQ I
Sbjct: 767 SEQQI 771
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2310**

A DNA sequence (GBSx2451) was identified in *S.agalactiae* <SEQ ID 7109> which encodes the amino acid sequence <SEQ ID 7110>. This protein is predicted to be phosphomethylpyrimidine kinase (thiD). Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2051(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC22074 GB:U32725 phosphomethylpyrimidine kinase (thiD)
          [Haemophilus influenzae Rd]
 Identities = 29/78 (37%), Positives = 48/78 (61%), Gaps = 2/78 (2%)

Query: 4  RNVLAISGNDIFSGGGLHADLATYVVNKLHGFVAVTCLTAMSDKG-FEVIPIEASILKQQ 62
          + VL I+G+D   G G+ ADL T+ +  + G  A+T +TA +  G F++ PI    ++ Q
Sbjct: 5  KQVLTIAGSDSGGGAGIQADLKTFQMRGVFGTSAITAVTAQNTLGVFDIHPIPLKTIQAQ 64

Query: 63 LESLK-DVEFGSIKLGLL 79
          LE++K D +  S K+G+L
Sbjct: 65 LEAVKNDFQIASCKIGML 82
```

There is also homology to SEQ ID 4408.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2311**

A DNA sequence (GBSx2452) was identified in *S.agalactiae* <SEQ ID 7111> which encodes the amino acid sequence <SEQ ID 7112>. Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -7.43   Transmembrane  109 - 125 ( 102 - 129)
   INTEGRAL    Likelihood = -1.28   Transmembrane   84 - 100 (  84 - 100)

----- Final Results -----
               bacterial membrane --- Certainty=0.3972(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA22372 GB:AL034446 putative transmembrane protein
          [Streptomyces coelicolor A3(2)]
 Identities = 25/93 (26%), Positives = 43/93 (45%), Gaps = 1/93 (1%)

Query: 62  SASVEILCRGWLLPVSATKYSKIVSVSISSIFFGLLHSANNHVSLISIFNLCL-FGLFLS 120
           +A+ E++ RG L  +        +++ ++ + FGL+H  N   +L      + +  G  L+
Sbjct: 143 AATEEVVFRGVLFRIIEEHIGTYLALGLTGLVFGLMHLLNEDATLWGALAIAIEAGFMLA 202
```

```
Query: 121 LYVILKGNIWGACGIHGAWNCVQGSVFGIEVSG 153
                   N+W   G+H  WN   G VF   VSG
Sbjct: 203 AAYAATRNLWLTIGVHFGWNFAAGGVFSTVVSG 235
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2312**

A DNA sequence (GBSx2453) was identified in *S.agalactiae* <SEQ ID 7113> which encodes the amino acid sequence <SEQ ID 7114>. This protein is predicted to be pppL protein. Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5796(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA10712 GB:AJ132604 pppL protein [Lactococcus lactis]
 Identities = 38/64 (59%), Positives = 51/64 (79%)

Query: 1  MEISLLTDIGQRRSNNQDFINQFENKAGVPLIILADGMGGHRAGNIASEMTVTDLGSDWA 60
          ME S+L+DIG +RS NQD++   + N+AG  L +LADGMGGH+AGN+AS++TV DLG  W+
Sbjct: 1  MEYSILSDIGSKRSTNQDYVGTYVNRAGYQLFLLADGMGGHKAGNVASKLTVEDLGKLWS 60

Query: 61 ETDF 64
          ET F
Sbjct: 61 ETFF 64
```

There is also homology to SEQ ID 3022:

```
Identities = 58/74 (78%), Positives = 69/74 (92%)

Query: 1  MEISLLTDIGQRRSNNQDFINQFENKAGVPLIILADGMGGHRAGNIASEMTVTDLGSDWA 60
          M+ISL TDIGQ+RSNNQDFIN+F+NK G+ L+ILADGMGGHRAGNIASEMTVTDLG +W
Sbjct: 1  MKISLKTDIGQKRSNNQDFINKFDNKKGITLVILADGMGGHRAGNIASEMTVTDLGREWV 60

Query: 61 ETDFSELSEIRDWM 74
          +TDF+ELS+IRDW+
Sbjct: 61 KTDFTELSQIRDWL 74
```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2313**

A DNA sequence (GBSx2454) was identified in *S.agalactiae* <SEQ ID 7115> which encodes the amino acid sequence <SEQ ID 7116>. This protein is predicted to be sunL protein. Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
```

```
               bacterial cytoplasm --- Certainty=0.1631(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA10711 GB:AJ132604 sunL protein [Lactococcus lactis]
 Identities = 48/81 (59%), Positives = 67/81 (82%)

Query: 1   MSILSSVCQTLRKGGIITYSTCTIFEEENFQVIEKFLENHPNFEQVELSHTQEDIVKRGC 60
            + IL+S  ++L+K GI+ YSTCTIF+EENF V+ +FLENHPNFEQVE+S+ + +++K GC
Sbjct: 342 LEILNSASKSLKKSGIMVYSTCTIFDEENFDVVHEFLENHPNFEQVEISNEKPEVIKEGC 401

Query: 61  ISISPEQYHTDGFFIGQVKRI 81
            + I+PE YHTDGFFI + K+I
Sbjct: 402 LFITPEMYHTDGFFIAKFKKI 422
```

There is also homology to SEQ ID 3018:

```
 Identities = 64/82 (78%), Positives = 74/82 (90%)

Query: 1   MSILSSVCQTLRKGGIITYSTCTIFEEENFQVIEKFLENHPNFEQVELSHTQEDIVKRGC 60
            + ILSSVCQTLRKGGIITYSTCTIF+EEN QVIE FL++HPNFEQV+L+HTQ DIVK G
Sbjct: 359 LEILSSVCQTLRKGGIITYSTCTIFDEENRQVIEAFLQSHPNFEQVKLNHTQADIVKDGY 418

Query: 61  ISISPEQYHTDGFFIGQVKRIL 82
            + I+PEQY TDGFFIGQV+R+L
Sbjct: 419 LIITPEQYQTDGFFIGQVRRVL 440
```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

## Example 2314

A DNA sequence (GBSx2455) was identified in *S.agalactiae* <SEQ ID 7117> which encodes the amino acid sequence <SEQ ID 7118>. This protein is predicted to be PTS permease for mannose subunit IIPMan. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -9.18   Transmembrane   32 -  48 (  30 -  58)
   INTEGRAL    Likelihood = -8.07   Transmembrane  127 - 143 ( 122 - 146)
   INTEGRAL    Likelihood = -2.07   Transmembrane   56 -  72 (  56 -  72)
   INTEGRAL    Likelihood = -1.44   Transmembrane   87 - 103 (  86 - 103)
   INTEGRAL    Likelihood = -0.53   Transmembrane  105 - 121 ( 105 - 121)

----- Final Results -----
                bacterial membrane --- Certainty=0.4673(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF81084 GB:AF228498 AgaW [Escherichia coli]
 Identities = 38/122 (31%), Positives = 68/122 (55%), Gaps = 7/122 (5%)

Query: 25  KVPETKSIIRLTALAFLVCSILVVELVSMRELISSISFIGILVGSGPVNSFVHHIPQNLM 84
            ++P T  +  L A  +L        L+++       +F+ I  G+    + +  +PQ L+
Sbjct: 126 RMPRTPILAALNACNYLA-------LLALGNFYFLCAFLPIYFGAEHAKTIIDVLPQRLI 178

Query: 85  NGLSAAGGLLPAVGFAMLMKLLWTNKLAVFYLLGFVLTAYLKLPAVAVAALGAVICVISS 144
            +GL  AGG++PA+GFA+L+K++  N    +++LGFV  A+LKLP +A+A     + +I
Sbjct: 179 DGLGVAGGIMPAIGFAVLLKIMMKNVYIPYFILGFVAAAWLKLPVLAIACPALAMALIDL 238

Query: 145 QR 146
```

```
              R
Sbjct: 239 LR 240
```

There is also homology to SEQ ID 1636:

```
Identities = 104/109 (95%), Positives = 108/109 (98%)

Query: 56  LISSISFIGILVGSGPVNSFVHHIPQNLMNGLSAAGGLLPAVGFAMLMKLLWTNKLAVFY 115
            +I+SISFIGILVGSGPVN+FV HIPQNLMNGLSAAGGLLPAVGFAMLMKLLWTNKLAVFY
Sbjct: 149 IIASISFIGILVGSGPVNAFVEHIPQNLMNGLSAAGGLLPAVGFAMLMKLLWTNKLAVFY 208

Query: 116 LLGFVLTAYLKLPAVAVAALGAVICVISSQRDIELDAITRGAISKQTTF 164
            LLGFVLTAYLKLPAVAVAALGAVICVISSQRD+ELDAITRGAISKQTTF
Sbjct: 209 LLGFVLTAYLKLPAVAVAALGAVICVISSQRDLELDAITRGAISKQTTF 257
```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2315

A DNA sequence (GBSx2456) was identified in *S.agalactiae* <SEQ ID 7119> which encodes the amino acid sequence <SEQ ID 7120>. Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -8.12    Transmembrane  121 - 137 ( 118 - 144)
   INTEGRAL    Likelihood = -5.52    Transmembrane   91 - 107 (  89 - 111)
   INTEGRAL    Likelihood = -5.20    Transmembrane  166 - 182 ( 162 - 192)

----- Final Results -----
              bacterial membrane --- Certainty=0.4248(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB15963 GB:Z99124 phosphotransferase system (PTS)
           beta-glucoside-specific enzyme IIABC component [Bacillus subtilis]
 Identities = 76/201 (37%), Positives = 122/201 (59%), Gaps = 3/201 (1%)

Query: 1   MIKALLALLLVFKILTPSSQTYILLNLFADGVFYFLPILIAITAAQKLKANPILALGTVV 60
            MIK L+AL + F  +   SQ +++L    DG FYFLP+L+A++AA+K  +NP +A
Sbjct: 121 MIKGLVALAVTFGWMAEKSQVHVILTAVGDGAFYFLPLLLAMSAARKFGSNPYVAAAIAA 180

Query: 61  MLLHPNWANLVASGKPVSLFHTIPFTLTNYASSVIPIILIICVQAYIEKYLKQIIPKSLR 120
             +LHP+   L+ +GKP+S F  +P T   Y+S+VIPI+L I + +Y+EK++ +    SL+
Sbjct: 181 AILHPDLTALLGAGKPIS-FIGLPVTAATYSSTVIPILLSIWIASYVEKWIDRFTHASLK 239

Query: 121 LVLVPMLIFLSMGILSFSILGPMGTIAGQYLAVIFTFLSKYASW-APAFLVGAFAPILIM 179
            L++VP    L +  L+   +GP+G I G+YL+    +L  +A   A  FL G F+ ++IM
Sbjct: 240 LIVVPTFTLLIVVPLTLITVGPLGAILGEYLSSGVNYLFDHAGLVAMIFLAGTFS-LIIM 298

Query: 180 FGVHSGIAALGITQLAKLGVD 200
             G+H     + I  +A+ G D
Sbjct: 299 TGMHYAFVPIMINNIAQNGHD 319
```

There is also homology to SEQ ID 2884.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2316**

A DNA sequence (GBSx2457) was identified in *S.agalactiae* <SEQ ID 7121> which encodes the amino acid sequence <SEQ ID 7122>. This protein is predicted to be glucose kinase. Analysis of this protein sequence reveals the following:

```
Possible site: 54
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1180(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14416 GB:Z99116 glucose kinase [Bacillus subtilis]
 Identities = 32/57 (56%), Positives = 41/57 (71%)

Query: 1   MVIGGGVSAAGEFLRSRVEKYFVTFAFPQVKKSTKIKIAELGNDAGIIGAASLANQQ 57
            +V+GGGVS AGE LRS+VEK F   AFP+  ++  I IA LGNDAG+IG A +A  +
Sbjct: 258 IVLGGGVSRAGELLRSKVEKTFRKCAFPRAAQAADISIAALGNDAGVIGGAWIAKNE 314
```

There is also homology to SEQ ID 198. An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 50/56 (89%), Positives = 53/56 (94%)

Query: 1   MVIGGGVSAAGEFLRSRVEKYFVTFAFPQVKKSTKIKIAELGNDAGIIGAASLANQ 56
            +VIGGGVSAAGEFLRSR+EKYFVTF FPQV+ STKIKIAELGNDAGIIGAASLA Q
Sbjct: 264 VVIGGGVSAAGEFLRSRIEKYFVTFTFPQVRYSTKIKIAELGNDAGIIGAASLARQ 319
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 2317**

A DNA sequence (GBSx2458) was identified in *S.agalactiae* <SEQ ID 7123> which encodes the amino acid sequence <SEQ ID 7124>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14385 GB:Z99116 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 37/86 (43%), Positives = 51/86 (59%)

Query: 3  MSVILIIVILLAFVAWASWNYWRVRRAAKFLDNESFQKEMSRGQLIDIREAGAFHRKHIL 62
           MS +++++I   AF+ +    +Y    +R   K L   E F+      + QLID+RE    F    HIL
Sbjct: 1  MSNMIVLIIFPAFIIYMIASYVYQQRIMKTLTEEEFRAGYRKAQLIDVREPNEFEGGHIL 60

Query: 63 GARNIPASQFKVALSALRKDKPVLLY 88
           GARNIP SQ K    + +R DKPV LY
Sbjct: 61 GARNIPLSQLKQRKNEIRTDKPVYLY 86
```

There is also homology to SEQ ID 202. An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 51/108 (47%), Positives = 70/108 (64%)
```

```
Query: 1   MDMSVILIIVILLAFVAWASWNYWRVRRAAKFLDNESFQKEMSRGQLIDIREAGAFHRKH 60
           M    +++ ++L+  V + +WNY+  R+ AK +DNE+F+  M +GQLID+RE  AF  KH
Sbjct: 1   MSPITLILWLLLVGIVGYYTWNYFSFRKMAKQVDNETFKDVMRQGQLIDLREPAAFRTKH 60

Query: 61  ILGARNIPASQFKVALSALRKDKPVLLYDASRGQSIPRIVLLLRKERF 108
           ILGARN PA QF  A+  LRKDKPVL+Y+  R Q     V  L+K  F
Sbjct: 61  ILGARNFPAQQFDAAIKGLRKDKPVLIYENMRPQYRVPAVKKLKKAGF 108
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 2318**

A DNA sequence (GBSx2459) was identified in *S.agalactiae* <SEQ ID 7125> which encodes the amino acid sequence <SEQ ID 7126>. This protein is predicted to be surface protein Rib. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1892(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2319**

A DNA sequence (GBSx2460) was identified in *S.agalactiae* <SEQ ID 7127> which encodes the amino acid sequence <SEQ ID 7128>. Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3522(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2320**

A DNA sequence (GBSx2461) was identified in *S.agalactiae* <SEQ ID 7129> which encodes the amino acid sequence <SEQ ID 7130>. Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2770(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```

```
                    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAB18708 GB:U38906 ORF33 [Bacteriophage r1t]
 Identities = 56/85 (65%), Positives = 66/85 (76%), Gaps = 1/85 (1%)

Query: 1  MTNFATTDDVILLWRQLSVDEIKRAEALLETVSDTLRLEASKVGKNLDEMILETP-YFAT 59
           M  FAT DD+ +LWR L  DE +RAE LLE VSD+LR EA KVG++L  MI E P YFA+
Sbjct: 1  MNPFATVDDLTMLWRPLKGDEKERAEKLLEIVSDSLREEADKVGRDLYAMIAEKPSYFAS 60

Query: 60 VLKSVTVDIVARTLMTATQGEPMSQ 84
           V+KSVTVDIVARTLMT+T  EPM+Q
Sbjct: 61 VVKSVTVDIVARTLMTSTDQEPMTQ 85
```

There is also homology to SEQ ID 1432.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2321**

A DNA sequence (GBSx2462) was identified in *S.agalactiae* <SEQ ID 7131> which encodes the amino acid sequence <SEQ ID 7132>. This protein is predicted to be regulatory protein TypA (typA). Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
                 bacterial cytoplasm --- Certainty=0.2238(Affirmative) < succ>
                  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB06351 GB:AP001516 GTP-binding protein TypA/BipA (tyrosine
           phosphorylated protein A) [Bacillus halodurans]
 Identities = 175/237 (73%), Positives = 204/237 (85%), Gaps = 1/237 (0%)

Query: 1   MEDIFVGETVTPTDAIEPLPVLRIDEPTLQMTFLVNNSPFAGREGKWITSRKVEERLLAE 60
            ME+I VGETV P D  +PLP+LRIDEPTLQMTFLVNNSPFAGREGK +TSRK+EERL AE
Sbjct: 281 MEEINVGETVCPVDHQDPLPILRIDEPTLQMTFLVNNSPFAGREGKHVTSRKLEERLRAE 340

Query: 61  LQTDVSLRVDPTDSPDKWTVSGRGELHLSILIETMRREGYELQVSRPEVIIKEIDGVQCE 120
            L+TDVSLRV+ TDSPD W VSGRGELHLSILIE MRREGYELQVS+PEVII+EIDGVQCE
Sbjct: 341 LETDVSLRVENTDSPDMWVVSGRGELHLSILIENMRREGYELQVSKPEVIIREIDGVQCE 400

Query: 121 PFERVQIDTPEEYQGAIIQSLSERKGDMLDMQMVGNGQTRLIFLIPARGLIGYSTEFLSM 180
            P ERVQID PEEY GA+++SL ERKG+ML+M   G+GQ RL F++PARGLIGY+TEFLS
Sbjct: 401 PVERVQIDVPEEYTGAVMESLGERKGEMLNMTNTGSGQVRLEFMVPARGLIGYTTEFLSQ 460

Query: 181 TRGYGIMNHTFDQYLPVVQGEIGGRHRGALVSIENGKATTYSIMRIEERGNLSFVNP 237
            TRGYGI+NH+FD Y PV  G++GGR +G LVS+E GKAT Y I+++E+RG + FV P
Sbjct: 461 TRGYGIINHSFDSYQPVTPGQVGGRRQGVLVSMETGKATQYGIIQVEDRGTI-FVEP 516
```

There is also homology to SEQ ID 206. An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 228/237 (96%), Positives = 233/237 (98%), Gaps = 1/237 (0%)

Query: 1   MEDIFVGETVTPTDAIEPLPVLRIDEPTLQMTFLVNNSPFAGREGKWITSRKVEERLLAE 60
            MEDIFVGET+TPTD +E LP+LRIDEPTLQMTFLVNNSPFAGREGKWITSRKVEERLLAE
Sbjct: 284 MEDIFVGETITPTDCVEALPILRIDEPTLQMTFLVNNSPFAGREGKWITSRKVEERLLAE 343

Query: 61  LQTDVSLRVDPTDSPDKWTVSGRGELHLSILIETMRREGYELQVSRPEVIIKEIDGVQCE 120
```

```
           LQTDVSLRVDPTDSPDKWTVSGRGELHLSILIETMRREGYELQVSRPEVIIKEIDGV+CE
Sbjct: 344 LQTDVSLRVDPTDSPDKWTVSGRGELHLSILIETMRREGYELQVSRPEVIIKEIDGVKCE 403

Query: 121 PFERVQIDTPEEYQGAIIQSLSERKGDMLDMQMVGNGQTRLIFLIPARGLIGYSTEFLSM 180
           PFERVQIDTPEEYQGAIIQSLSERKGDMLDMQMVGNGQTRLIFLIPARGLIGYSTEFLSM
Sbjct: 404 PFERVQIDTPEEYQGAIIQSLSERKGDMLDMQMVGNGQTRLIFLIPARGLIGYSTEFLSM 463

Query: 181 TRGYGIMNHTFDQYLPVVQGEIGGRHRGALVSIENGKATTYSIMRIEERGNLSFVNP 237
           TRGYGIMNHTFDQYLPVVQGEIGGRHRGALVSIENGKATTYSIMRIEERG + FVNP
Sbjct: 464 TRGYGIMNHTFDQYLPVVQGEIGGRHRGALVSIENGKATTYSIMRIEERGTI-FVNP 519
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 2322**

A DNA sequence (GBSx2464) was identified in *S.agalactiae* <SEQ ID 7133> which encodes the amino acid sequence <SEQ ID 7134>. This protein is predicted to be pseudouridine synthase family 1 protein (rluB). Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1950(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14248 GB:Z99116 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 59/105 (56%), Positives = 85/105 (80%)

Query: 5   VKERIYPVGRLDWDTTGLLILTNDGDFTDKMIHPRNEIDKVYLARVKGIATKENLRPLTR 64
           + +RIYP+GRLD+DT+GLL+LTNDG+F +K++HP+ EIDK Y+A+VKGI  KE LR L R
Sbjct: 91  IPQRIYPIGRLDYDTSGLLLLTNDGEFANKLMHPKYEIDKTYVAKVKGIPPKELLRKLER 150

Query: 65  GVVIDGKKTKPARYTIIKVDHEKNRSVVELTIHEGRNHQVKKMFE 109
           G+ ++  KT PA+  ++ +D +K  S+++LTIHEGRN QV++MFE
Sbjct: 151 GIRLEEGKTAPAKAKLLSLDKKKQTSIIQLTIHEGRNRQVRRMFE 195
```

There is also homology to SEQ ID 4728:

```
Identities = 96/109 (88%), Positives = 106/109 (97%)

Query: 1   MLPQVKERIYPVGRLDWDTTGLLILTNDGDFTDKMIHPRNEIDKVYLARVKGIATKENLR 60
           +LPQVKERIYPVGRLDWDT+G+LILTNDGDFTD MIHPRNEIDKVYLARVKGIATKENLR
Sbjct: 94  LLPQVKERIYPVGRLDWDTSGVLILTNDGDFTDTMIHPRNEIDKVYLARVKGIATKENLR 153

Query: 61  PLTRGVVIDGKKTKPARYTIIKVDHEKNRSVVELTIHEGRNHQVKKMFE 109
           PLTRG+VIDGKKTKPARY I++V+ +K+RS+VELTIHEGRNHQVKKMFE
Sbjct: 154 PLTRGIVIDGKKTKPARYNIVRVEADKSRSIVELTIHEGRNHQVKKMFE 202
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 2323**

A DNA sequence (GBSx2466) was identified in *S.agalactiae* <SEQ ID 7135> which encodes the amino acid sequence <SEQ ID 7136>. This protein is predicted to be L-ribulose 5-phosphate 4-epimerase. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2827(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD45716 GB:AF160811 L-ribulose 5-phosphate 4-epimerase
           [Bacillus stearothermophilus]
 Identities = 68/103 (66%), Positives = 82/103 (79%)

Query: 2   QEMRERVCEANKSLPVHSLVKFTWGNVSEVDREAGLIVIKPSGVDYDQLTPENMVVTDLE 61
           +E+++ V EAN  LP + LV FTWGNVS +DRE GL+VIKPSGV YD+LT ++MVV DL
Sbjct: 3   EELKQAVLEANLQLPQYRLVTFTWGNVSGIDRERGLVVIKPSGVAYDKLTIDDMVVVDLT 62

Query: 62  GNIVEGDLNPSSDLPTHVQLYKAWPEVGGIVHTHSTEAVGWAQ 104
           GN+VEGDL PSSD PTH+ LYK +P +GGIVHTHST A  WAQ
Sbjct: 63  GNVVEGDLKPSSDTPTHLWLYKQFPGIGGIVHTHSTWATVWAQ 105
```

There is also homology to SEQ ID 4600:

```
Identities = 93/103 (90%), Positives = 96/103 (92%)

Query: 2   QEMRERVCEANKSLPVHSLVKFTWGNVSEVDREAGLIVIKPSGVDYDQLTPENMVVTDLE 61
           QEMRERVC ANKSLP H LVKFTWGNVSEV RE G IVIKPSGVDYD LTPENMVVTDL+
Sbjct: 6   QEMRERVCAANKSLPQHGLVKFTWGNVSEVCRELGRIVIKPSGVDYDLLTPENMVVTDLD 65

Query: 62  GNIVEGDLNPSSDLPTHVQLYKAWPEVGGIVHTHSTEAVGWAQ 104
           GN+VEGDLNPSSDLPTHV+LYKAWPEVGGIVHTHSTEAVGWAQ
Sbjct: 66  GNVVEGDLNPSSDLPTHVELYKAWPEVGGIVHTHSTEAVGWAQ 108
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 2324**

A DNA sequence (GBSx2467) was identified in *S.agalactiae* <SEQ ID 7137> which encodes the amino acid sequence <SEQ ID 7138>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3452(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAG05712 GB:AE004658 hypothetical protein [Pseudomonas aeruginosa]
 Identities = 141/200 (70%), Positives = 162/200 (80%), Gaps = 1/200 (0%)

Query: 10  LSLGTDYETLANRFRPIFREISAGNVEREKARALPYEPIEWLKKAGFGAVRVPSEYGGAG 69
           LS G DYE LA RFRPIF  I+ G VERE+ R LP+E I WLK+AGFGAVRVP E+GGAG
Sbjct: 14  LSEGADYELLAQRFRPIFARIAEGAVERERQRELPHEAIAWLKQAGFGAVRVPREHGGAG 73

Query: 70  ASIGQLFQLLIELAEADSNIPQALRAHFAFVEDRLNAPPGVDRDTWFARFVAGDLVGNGW 129
           AS+ QL QLLIELAEADSNI QALR HFAFVEDRLNA PG  RD W  RFV GDLVG  W
Sbjct: 74  ASLPQLVQLLIELAEADSNITQALRGHFAFVEDRLNAEPGPGRDRWLRRFVEGDLVGCAW 133

Query: 130 TEVGTVKIGDVITKVSAQGDG-FVLNGTKFYSTGSIFADWIDVYAQRADNGADVIAVVNA 188
           TEVG+V++G+V+T+VS + DG +V+NG+K+YSTGS+F+DWID+YAQR D GADVIA +
```

```
Sbjct: 134 TEVGSVRLGEVLTRVSRKDDGRWVVNGSKYYSTGSLFSDWIDLYAQRDDTGADVIAAIRT 193

Query: 189 RHAGVRHSDDWDGFGQRTTG 208
              GVR SDDWDGFGQRTTG
Sbjct: 194 DQPGVRQSDDWDGFGQRTTG 213
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2325**

A DNA sequence (GBSx2468) was identified in *S.agalactiae* <SEQ ID 7139> which encodes the amino acid sequence <SEQ ID 7140>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1919(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2326**

A DNA sequence (GBSx2474) was identified in *S.agalactiae* <SEQ ID 7141> which encodes the amino acid sequence <SEQ ID 7142>. Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2978(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2327**

A DNA sequence (GBSx2476) was identified in *S.agalactiae* <SEQ ID 7143> which encodes the amino acid sequence <SEQ ID 7144>. Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5402(Affirmative) < succ>
```

```
                 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2328**

A DNA sequence (GBSx2477) was identified in *S.agalactiae* <SEQ ID 7145> which encodes the amino acid sequence <SEQ ID 7146>. This protein is predicted to be mercuric reductase. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.2755(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA70224 GB:Y09024 mercuric reductase [Bacillus cereus]
 Identities = 190/247 (76%), Positives = 225/247 (90%)

Query: 1   MELGQLFHHLGSEITLMQRSERLLKEYDPEISESVEKALIEQGINLVKGATFERVEQSGE 60
           MELGQLFH+LGSE+TL+QRSERLLKEYDPEISESVEK+L+EQGINLVKGAT+ER+EQ+G+
Sbjct: 262 MELGQLFHNLGSEVTLIQRSERLLKEYDPEISESVEKSLVEQGINLVKGATYERIEQNGD 321

Query: 61  IKRVYVTVNGSREVIESDQLLVATGRKPNTDSLNLSAAGVETGKNNEILINDFGQTSNEK 120
           IK+V+V VNG + +IE+DQLLVATGR PNT +LNL AAGVE G   EI+I+D+ +T+N +
Sbjct: 322 IKKVHVEVNGKKRIIEADQLLVATGRTPNTATLNLRAAGVEIGSRGEIIIDDYSRTTNTR 381

Query: 121 IYAAGDVTLGPQFVYVVAAYEGGIITDNAIGGLNKKIDLSVVPAVTFTNPTVATVGLTEEQ 180
           IYAAGDVTLGPQFVYVVAAY+GG+   NAIGGLNKK++L VVP VTFT P +ATVGLTE+Q
Sbjct: 382 IYAAGDVTLGPQFVYVVAAYQGGVAAPNAIGGLNKKLNLEVVPGVTFTAPAIATVGLTEQQ 441

Query: 181 AKEKGYDVKTSVLPLGAVPRAIVNRETTGVFKLVADAETLKVLGVHIVSENAGDVIYAAS 240
           AKE GY+VKTSVLPL AVPRA+VNRETTGVFKLVAD++T+KVLG H+V+ENAGDVIYAA+
Sbjct: 442 AKENGYEVKTSVLPLDAVPRALVNRETTGVFKLVADSKTMKVLGAHVVAENAGDVIYAAT 501

Query: 241 LAVKFGL 247
           LAVKFGL
Sbjct: 502 LAVKFGL 508
```

There is also homology to SEQ ID 1820.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2329**

A DNA sequence (GBSx2478) was identified in *S.agalactiae* <SEQ ID 7147> which encodes the amino acid sequence <SEQ ID 7148>. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
```

```
                        bacterial cytoplasm --- Certainty=0.3642(Affirmative) < succ>
                         bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2330**

A DNA sequence (GBSx2479) was identified in *S.agalactiae* <SEQ ID 7149> which encodes the amino acid sequence <SEQ ID 7150>. This protein is predicted to be surface protein Rib. Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1936(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2331**

A DNA sequence (GBSx2480) was identified in *S.agalactiae* <SEQ ID 7151> which encodes the amino acid sequence <SEQ ID 7152>. This protein is predicted to be Nra. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1510(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9383> which encodes amino acid sequence <SEQ ID 9384> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 7153> which encodes the amino acid sequence <SEQ ID 7154>. Analysis of this protein sequence reveals the following:

```
     Possible site: 16
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.64   Transmembrane   22 -  38 (  22 -  38)

----- Final Results -----
               bacterial membrane --- Certainty=0.1256(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 42/157 (26%), Positives = 78/157 (48%), Gaps = 2/157 (1%)

Query: 71  LLGREFIDSQHFKDINAYFLRHFICYCYYFIPDFYFLNTSRLSY--SKDLYHLLDKGLAD 128
           LLG   ++S  FK I   F R FI      +PD +  +  R     +K  Y+ L   + +
Sbjct: 8   LLGNNILNSLPFKRILVSFSRLFISNLQVLLPDIHLFHYLRRQQKRNKSFYNTLKTIVEE 67

Query: 129 IFNLKGGNLTFSKHETVLLTMQLSNLIETFLAPLSVYVISSSNIRLQTYQVMLNQYFTSK 188
             + +G       +  +L T+QL  L++T+L P+ VY+++++   L      L+ YF
Sbjct: 68  WMSAEGIVGKLPSYHLLLFTIQLEELLKTYLPPIPVYLLTNNTAALDLMTNALSIYFPPA 127

Query: 189 IAEFFFVNYQTTQIDEKLLKKADIIIAERRYISSLKN 225
           IA    VN +     + + +K  +IIA+R+Y++ +++
Sbjct: 128 IATVMPVNVEIIPFKDIVKEKQSVIIADRQYLNLIQH 164
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 2332**

A DNA sequence (GBSx2481) was identified in *S.agalactiae* <SEQ ID 7155> which encodes the amino acid sequence <SEQ ID 7156>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1383(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2333**

A DNA sequence (GBSx2482) was identified in *S.agalactiae* <SEQ ID 7157> which encodes the amino acid sequence <SEQ ID 7158>. Analysis of this protein sequence reveals the following:

```
Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4145(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2334**

A DNA sequence (GBSx2484) was identified in *S.agalactiae* <SEQ ID 7159> which encodes the amino acid sequence <SEQ ID 7160>. Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.02    Transmembrane   34 -  50 (  34 -  50)

----- Final Results -----
               bacterial membrane --- Certainty=0.1808(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2335**

A DNA sequence (GBSx2485) was identified in *S.agalactiae* <SEQ ID 7161> which encodes the amino acid sequence <SEQ ID 7162>. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3488(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB52002 GB:AL109663 hypothetical protein [Streptomyces
           coelicolor A3(2)]
 Identities = 61/141 (43%), Positives = 86/141 (60%), Gaps = 2/141 (1%)

Query: 3   TYFDNFLKTNQAYADLHGTAHLPIKPKTKVAIVTCMDSRLHVAQALGLALGDAHILRNAG 62
           T  D ++  N+ YA       +  +P  +VA+V CMD+RL +  ALGL LGD H +RNAG
Sbjct: 5   TVTDRLVEANERYAAAFADPGMDARPVQRVAVVACMDARLDLHAALGLKLGDCHTIRNAG 64

Query: 63  GRVTDDVLRSLVISQQQLGTREIVVLHHTDCGAQTFTNEAFAAQLQRDLGVDMHGHDFLP 122
           G VTDDV+RSL ISQ+ LGTR + ++HHT CG +T T E F   L+ ++G
Sbjct: 65  GVVTDDVIRSLTISQRALGTRSVALIHHTGCGMETITEE-FRHDLELEVG-QRPAWAVEA 122

Query: 123 FNDIEESVREDVAKLHASPFL 143
           F D ++ VR+ + ++  SPFL
Sbjct: 123 FRDADQDVRQSIERVRTSPFL 143
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6469> which encodes the amino acid sequence <SEQ ID 6470>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2295(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 109/146 (74%), Positives = 128/146 (87%)

Query: 1   MTTYFDNFLKTNQAYADLHGTAHLPIKPKTKVAIVTCMDSRLHVAQALGLALGDAHILRN 60
            + +YF++F+  NQAY  LHGTAHLP+KPKTKVAIVTCMDSRLHVAQALGLALGDAHILRN
Sbjct: 1   LMSYFEHFMAANQAYVALHGTAHLPLKPKTKVAIVTCMDSRLHVAQALGLALGDAHILRN 60

Query: 61  AGGRVTDDVLRSLVISQQQLGTREIVVLHHTDCGAQTFTNEAFAAQLQRDLGVDMHGHDF 120
            AGGRVT+D++RSLVISQQQ+GTREIVVLHHTDCGAQTFTNE FA  +   LGVD+ G DF
Sbjct: 61  AGGRVTEDMIRSLVISQQQMGTREIVVLHHTDCGAQTFTNEGFAKHIHEHLGVDVSGQDF 120

Query: 121 LPFNDIEESVREDVAKLHASPFLREE 146
            LPF D+E+SVRED+AK+ AS  + ++
Sbjct: 121 LPFQDVEDSVREDMAKIRASSLISDD 146
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 2336**

A DNA sequence (GBSx2486) was identified in *S.agalactiae* <SEQ ID 7163> which encodes the amino acid sequence <SEQ ID 7164>. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0932(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAG08811 GB:AE004955 phosphoribosylaminoimidazole carboxylase,
          catalytic subunit [Pseudomonas aeruginosa]
 Identities = 20/27 (74%), Positives = 26/27 (96%)

Query: 1  MFKHAEEARGRGIKIIIAGAGGAAHLP 27
          +F++AEEA GRG+++IIAGAGGAAHLP
Sbjct: 46 LFQYAEEAEGRGLEVIIAGAGGAAHLP 72
```

There is also homology to SEQ ID 910:

```
Identities = 27/27 (100%), Positives = 27/27 (100%)

Query: 1   MFKHAEEARGRGIKIIIAGAGGAAHLP 27
            MFKHAEEARGRGIKIIIAGAGGAAHLP
Sbjct: 87  MFKHAEEARGRGIKIIIAGAGGAAHLP 113
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 2337**

A DNA sequence (GBSx2488) was identified in *S.agalactiae* <SEQ ID 7165> which encodes the amino acid sequence <SEQ ID 7166>. Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -6.85    Transmembrane    58 -  74 (  53 -  80)
   INTEGRAL    Likelihood = -5.79    Transmembrane   103 - 119 ( 101 - 122)

----- Final Results -----
               bacterial membrane --- Certainty=0.3739(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

```
                  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

There is also homology to SEQ IDs 880 and 9278.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2338**

A DNA sequence (GBSx2489) was identified in *S.agalactiae* <SEQ ID 7167> which encodes the amino acid sequence <SEQ ID 7168>. This protein is predicted to be short chain alcohol dehydrogenase. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1742(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9357> which encodes amino acid sequence <SEQ ID 9358> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD06605 GB:AE001530 putative oxidoreductase [Helicobacter
           pylori J99]
 Identities = 68/94 (72%), Positives = 79/94 (83%)

Query: 4   IDLLVNNAGLALGLDKSYEADFGDWMTMINTNVVGLIYLTRCILPKMVEVNRGLIINLGS 63
           ID L+NNAGLALGL+K+YE +  DW  MI+TN+ GL++LTR ILP M+E ++G IINLGS
Sbjct: 76  IDALINNAGLALGLNKAYECELDDWEVMIDTNIKGLLHLTRLILPSMIEHDQGTIINLGS 135

Query: 64  XAGTIPYPGANVYGASKAFVKQFSLNLRADLAGT 97
            AGT  YPG NVYGASKAFVKQFSLNLRADLAGT
Sbjct: 136 IAGTYAYPGGNVYGASKAFVKQFSLNLRADLAGT 169
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 7169> which encodes the amino acid sequence <SEQ ID 7170>. Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related sequence was also identified in GAS <SEQ ID 9121> which encodes the amino acid sequence <SEQ ID 9122>. Analysis of this protein sequence reveals the following:

```
     Possible site: 12
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 78/96 (81%), Positives = 87/96 (90%)

Query: 2   QSIDLLVNNAGLALGLDKSYEADFGDWMTMINTNVVGLIYLTRCILPKMVEVNRGLIINL 61
            Q I +LVNNAGLALGLDK+YEADF +WMTMINTN+VGLIYLTR +LP MV  + G+IINL
Sbjct: 82  QDITILVNNAGLALGLDKAYEADFENWMTMINTNIVGLIYLTRQLLPHMVSKDDGIIINL 141

Query: 62  GSXAGTIPYPGANVYGASKAFVKQFSLNLRADLAGT 97
            GS AGTIPYPGAN+YGASKAFVKQFSLNLRADLAG+
Sbjct: 142 GSTAGTIPYPGANIYGASKAFVKQFSLNLRADLAGS 177
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 2339**

A DNA sequence (GBSx2492) was identified in *S.agalactiae* <SEQ ID 7171> which encodes the amino acid sequence <SEQ ID 7172>. This protein is predicted to be mercuric reductase. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2115(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAC14663 GB:Y10855 mercuric reductase [Bacillus licheniformis]
 Identities = 68/104 (65%), Positives = 82/104 (78%)

Query: 1   MNKFKVNISGMTCTGCEKHVESALEKIGAKNIESSYRRGEAVFELPDDIEVESAIKAIDE 60
            M K++VN+ GMTCTGCE+HV  ALE +GAK IE  YRRGEAVFELP+ +EVE+A KAI E
Sbjct: 1   MKKYRVNVQGMTCTGCEEHVAVALENMGAKRIEVDYRRGEAVFELPNGLEVETAKKAIAE 60

Query: 61  ANYQAGEIEEVSSLENVALINEDNYDLLIIGSGAAAFSSAIKAI 104
            A YQ GE EEV S E + L +E +YD +IIGSG AAFSSAI+A+
Sbjct: 61  AKYQPGEAEEVQSQELIQLGDEGDYDYIIIGSGGAAFSSAIEAV 104
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2340**

A DNA sequence (GBSx2494) was identified in *S.agalactiae* <SEQ ID 7173> which encodes the amino acid sequence <SEQ ID 7174>. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3341(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2341**

A DNA sequence (GBSx2495) was identified in *S.agalactiae* <SEQ ID 7175> which encodes the amino acid sequence <SEQ ID 7176>. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4989(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2342**

A DNA sequence (GBSx2496) was identified in *S.agalactiae* <SEQ ID 7177> which encodes the amino acid sequence <SEQ ID 7178>. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2569(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2343**

A DNA sequence (GBSx2497) was identified in *S.agalactiae* <SEQ ID 7179> which encodes the amino acid sequence <SEQ ID 7180>. This protein is predicted to be DNA polymerase III alpha subunit (dnaE). Analysis of this protein sequence reveals the following:

```
Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3124(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4095> which encodes the amino acid sequence <SEQ ID 4096>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
```

```
              bacterial cytoplasm --- Certainty=0.2600(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 186/237 (78%), Positives = 214/237 (89%)

Query: 10  DPVKHNLIFERFLNEERYSMPDIDIDLPDIYRGEFLRYVRNRYGSMHSAQIVTFSTFGAK 69
            DPV+H+L+FERFLN+ERYSMPDIDIDLPDIYR EFLRYVRNRYGS HSAQIVTFSTFG K
Sbjct: 321 DPVQHDLLFERFLNKERYSMPDIDIDLPDIYRSEFLRYVRNRYGSDHSAQIVTFSTFGPK 380

Query: 70  QAIRDVFKRFGASEYELTNITKKIHFRDNLTSVYNRNLAFRQIIDSKIEYQKAYDIAKRI 129
            QAIRDVFKRFG  EYELTN+TKKI F+D+L +VY ++++FRQ+I+S+ E+QKA+ IAKRI
Sbjct: 381 QAIRDVFKRFGVPEYELTNLTKKIGFKDSLATVYEKSISFRQVINSRTEFQKAFAIAKRI 440

Query: 130 EGNPRQTSIHAAGVVMSDDLLTDHIPLKNGEDMMITQYDASSVEDNGLLKMDFLGLRNLT 189
            EGNPRQTSIHAAG+VMSDD LT+HIPLK+G+DMMITQYDA +VE NGLLKMDFLGLRNLT
Sbjct: 441 EGNPRQTSIHAAGIVMSDDALTNHIPLKSGDDMMITQYDAHAVEANGLLKMDFLGLRNLT 500

Query: 190 FVQKMKEKVDKDYGISIQLETIDLEDKETLKLFAAGQTKGIFQFEQSGAINLLRRIR 246
            FVQKM+EKV KDYG  I +  IDLED +TL LFA G TKGIFQFEQ+GAINLL+RI+
Sbjct: 501 FVQKMQEKVAKDYGCQIDITAIDLEDPQTLALFAKGDTKGIFQFEQNGAINLLKRIK 557
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 2344**

A DNA sequence (GBSx2498) was identified in *S.agalactiae* <SEQ ID 7181> which encodes the amino acid sequence <SEQ ID 7182>. This protein is predicted to be a methylase. Analysis of this protein sequence reveals the following:

```
Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2121(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAG21729 GB:AF116907 putative methylase [Corynebacterium hoagii]
 Identities = 48/160 (30%), Positives = 85/160 (53%), Gaps = 6/160 (3%)

Query: 97   EPDDSENGHNDTDLEETDNQIPEEEVVETIPEIPVTDFYFPEDLTDFYPKTARDKVETNI 156
            EP+        +  E  +  + ++E       +P TDF    D+    P  A+ +V  NI
Sbjct: 1236 EPEAPTQPEAASAAETAEPAVEQQEPRAGPQSVPATDFALGTDV--HVPSGAKARVRANI 1293

Query: 157  VAIRLVKNLEVEHRNASPSEQELLAKYVGWGGLANEFFDD---YNPKFSKEREELKSLVT 213
             A RLV  L+ + R A+  EQ +LA++ GWG +  E FD+   +  +++ ER  L  L+
Sbjct: 1294 AAARLVLELDEQQRPATAEEQAVLAQWSGWGAVP-EVFDNRSKFLSEWADERAALLDLLG 1352

Query: 214  DKEYSDMKQSSLTAYYTDPSLIRQMWGIVERDGFTGWQIL 253
            +K +S  ++++L A+YTDP+++ ++W  V+R G     +L
Sbjct: 1353 EKGFSQARETTLNAHYTDPAIVGELWRAVQRAGLPDGALL 1392
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2345**

A DNA sequence (GBSx2499) was identified in *S.agalactiae* <SEQ ID 7183> which encodes the amino acid sequence <SEQ ID 7184>. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1111(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2346**

A DNA sequence (GBSx2501) was identified in *S.agalactiae* <SEQ ID 7185> which encodes the amino acid sequence <SEQ ID 7186>. Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4752(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA61516 GB:X89232 DNA-directed RNA polymerase [Pediococcus
          acidilactici]
 Identities = 48/53 (90%), Positives = 52/53 (97%)

Query: 5  KKPETINYRTLKPEREGLFDEVIFGPTKDWECACGKYKRIRYKGIICDRCGVE 57
           KKPETINYRTLKPE++GLFDE IFGPTKD+ECACGKYKRIRYKGI+CDRCGVE
Sbjct: 29 KKPETINYRTLKPEKDGLFDERIFGPTKDYECACGKYKRIRYKGIVCDRCGVE 81
```

There is also homology to SEQ ID 384.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 2347**

A DNA sequence (GBSx2502) was identified in *S.agalactiae* <SEQ ID 7187> which encodes the amino acid sequence <SEQ ID 7188>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3080(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC00282 GB:AF008220 YtlR [Bacillus subtilis]
 Identities = 61/216 (28%), Positives = 98/216 (45%), Gaps = 28/216 (12%)

Query: 8    IPCTYYPVGSGNDFARALKIPNL---------KETLTAIQTERLKEINCFIYDKGLIL-- 56
            I  ++ P G+ NDF+R   I  +         K  LT  +T  L  +N F+ DK  IL
Sbjct: 86   IELSFVPAGAYNDFSRGFSIKKIDLIQEIKKVKRPLT--RTFHLGSVN-FLQDKSQILYF 142

Query: 57   -NSLDLGFAAYVVWKASNSKIKNILNRYRLGKITYIVIAIKSLLHSSK------VQVLVE 109
             N + +GF AYV  KA    ++ +    RL  + Y +    S LH+S       +    E
Sbjct: 143 MNHIGIGFDAYVNKKAMEFPLRRVFLFLRLRFLVYPL----SHLHASATFKPFTLACTTE 198

Query: 110 GETGQQIKLNDLYFFALANNTYFGGGITIWPKASALTAELDMVYAKGHTFLKRLSILLSL 169
             ET +    +D++F  ++N+ ++GGG+   P A+      D+V  +   FLK+  +L  +
Sbjct: 199 DETRE---FHDVWFAVVSNHPFYGGGMKAAPLANPREKTFDIVIVENQPFLKKYWLLCLM 255

Query: 170 VFKRHTTSKSIKHQTFKAMTVYFPKNSLIEIDGEIV 205
             F +HT    +     K +T Y         DGEI+
Sbjct: 256 AFGKHTKMDGVTMFKAKDITFYTKDKIPFHADGEIM 291
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2348**

A DNA sequence (GBSx2503) was identified in *S.agalactiae* <SEQ ID 7189> which encodes the amino acid sequence <SEQ ID 7190>. This protein is predicted to be protease subunit HflC (hflC). Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1809(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAG08326 GB:AE004907 protease subunit HflC [Pseudomonas aeruginosa]
 Identities = 182/202 (90%), Positives = 194/202 (95%)

Query: 1    MSQTERAVLLQFGKVVQTDVKPGLHVKVPYVNQVRKFDGRLLTLDAPTQRFLTLEKKAVM 60
            + QTERAV+L+FG+VV++DVKPGLH K+PYVNQVRKFD RLLTLDAPTQRFLTLEKKAVM
Sbjct: 26   VQQTERAVMLRFGRVVESDVKPGLHFKIPYVNQVRKFDARLLTLDAPTQRFLTLEKKAVM 85

Query: 61   VDAYAKWRVKDAERFYTATSGLKQIADERLSRRLESGLRDQFGKRTLHEVVSGERDALMA 120
            VDAYAKWRV DAERFYTATSGLKQIADERLSRRLE+GLRDQFGKRTLHEVVSGERDALM
Sbjct: 86   VDAYAKWRVADAERFYTATSGLKQIADERLSRRLEAGLRDQFGKRTLHEVVSGERDALMG 145

Query: 121 DITGSLNRMAEKELGIEVLDVRVKAIDLPKEVNRSVFERMSTEREREAREHRAKGNELGE 180
            DIT SLNRMA+KELGIEV+DVRVKAIDLPKEVNRSVFERMSTEREREAREHRAKG EL E
Sbjct: 146 DITASLNRMAQKELGIEVIDVRVKAIDLPKEVNRSVFERMSTEREREAREHRAKGRELAE 205

Query: 181 GIRADADRQRRVLLAEAYRESE 202
            GIRADADRQRRV++AEAYRESE
Sbjct: 206 GIRADADRQRRVIVAEAYRESE 227
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2349**

A DNA sequence (GBSx2504) was identified in *S.agalactiae* <SEQ ID 7191> which encodes the amino acid sequence <SEQ ID 7192>. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2350**

A DNA sequence (GBSx2505) was identified in *S.agalactiae* <SEQ ID 7193> which encodes the amino acid sequence <SEQ ID 7194>. This protein is predicted to be ABC transporter (ATP-binding; daunorubicin resistance). Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1846(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB15892 GB:Z99123 similar to ABC transporter (ATP-binding
           protein) [Bacillus subtilis]
 Identities = 88/231 (38%), Positives = 132/231 (57%), Gaps = 13/231 (5%)

Query: 10  QVIGYLPDVPKFYDYMTAQEYLQLC---AGLAQNKTSLPIADLLEQVGLADN-QQRISTY 65
             ++IGYLP  P FY +MTA E+L      +GL++ K    I ++LE VGL +   +RI  Y
Sbjct: 69  RLIGYLPQYPAFYSWMTANEFLTFAGRLSGLSKRKCQEKIGEMLEFVGLHEAAHKRIGGY 128

Query: 66  SRGMKQRLGLAQALIHXXKILICDEPTSALDPQGRQEILSIISQLRGQKTVIFSTHILSD 125
            S GMKQRLGLAQAL+H  K LI DEP SALDP GR E+L ++ +L+    V+FSTH+L D
Sbjct: 129 SGGMKQRLGLAQALLHKPKFLILDEPVSALDPTGRFEVLDMMRELKKHMAVLFSTHVLHD 188

Query: 126 VEKVCDQVLILTKSGIH---NLEDLRDKASASVNQLNLLIKVSDNEAQKLALRFPLNQKD 182
             E+VCDQV+I+    I     L++L+ +   +V  L++  K+     +K  +   + +
Sbjct: 189 AEQVCDQVVIMKNGEISWKGELQELKQQQQTNVFTLSVKEKLEGWLEEKPYVSAIVYKNP 248

Query: 183 QYYKVHLELSEANNREQALASFYRYLVEQEITPYFIELLEDSLEDFYLEVI 233
               +   EL + +     L+      + + +T    E   +SLED YL+V+
Sbjct: 249 S--QAVFELPDIHAGRSLLSD----CIRKGLTVTRFEQKTESLEDVYLKVV 293
```

There is also homology to SEQ ID 686.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2351**

A DNA sequence (GBSx2506) was identified in *S.agalactiae* <SEQ ID 7195> which encodes the amino acid sequence <SEQ ID 7196>. Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0679(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with glycine-rich cell wall proteins (*e.g.* GB:AL161589 – the glycine-rich cell wall protein from *Arabidopsis thaliania*) and to SEQ ID 6882.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2352**

A DNA sequence (GBSx2507) was identified in *S.agalactiae* <SEQ ID 7197> which encodes the amino acid sequence <SEQ ID 7198>. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2890(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2353**

A DNA sequence (GBSx2508) was identified in *S.agalactiae* <SEQ ID 7199> which encodes the amino acid sequence <SEQ ID 7200>. Analysis of this protein sequence reveals the following:

```
Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2410(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9329> which encodes amino acid sequence <SEQ ID 9330> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 9330 (GBS678) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 163 (lane 18; MW 53kDa), Figure 164 (lane 2 & 3; MW 53kDa) and Figure 188 (lane 7; MW 53kDa). Purified protein is shown in Figure 242, lanes 6 & 7.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2354**

A DNA sequence (GBSx2509) was identified in *S.agalactiae* <SEQ ID 7201> which encodes the amino acid sequence <SEQ ID 7202>. This protein is predicted to be surface protein Rib. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.2025(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2355**

A DNA sequence (GBSx2510) was identified in *S.agalactiae* <SEQ ID 7203> which encodes the amino acid sequence <SEQ ID 7204>. This protein is predicted to be surface protein Rib. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1892(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2356**

A DNA sequence (GBSx2511) was identified in *S.agalactiae* <SEQ ID 7205> which encodes the amino acid sequence <SEQ ID 7206>. This protein is predicted to be surface protein Rib. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1892(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2357**

A DNA sequence (GBSx2512) was identified in *S.agalactiae* <SEQ ID 7207> which encodes the amino acid sequence <SEQ ID 7208>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0999(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2358**

A DNA sequence (GBSx2514) was identified in *S.agalactiae* <SEQ ID 7209> which encodes the amino acid sequence <SEQ ID 7210>. This protein is predicted to be surface protein Rib. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1892(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2359**

A DNA sequence (GBSx2515) was identified in *S.agalactiae* <SEQ ID 7211> which encodes the amino acid sequence <SEQ ID 7212>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2041(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2360**

A DNA sequence (GBSx2516) was identified in *S.agalactiae* <SEQ ID 7213> which encodes the amino acid sequence <SEQ ID 7214>. This protein is predicted to be 30S ribosomal protein S6 (rpsF). Analysis of this protein sequence reveals the following:

```
Possible site: 51
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3607(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9423> which encodes amino acid sequence <SEQ ID 9424> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB16128 GB:Z99124 ribosomal protein S6 (BS9) [Bacillus subtilis]
 Identities = 41/72 (56%), Positives = 58/72 (79%), Gaps = 1/72 (1%)

Query: 1  MVARFDSILSDNGATVVESKDWEKRRLAYEIQDFTEGLYHIVNVEAEDAVALNEFDRLSK 60
          ++ RF+++L+ NGA +  +KDW KRRLAYEI DF +G Y IVNV++ DA A+ EFDRL+K
Sbjct: 22 VIERFNNVLTSNGAEITGTKDWGKRRLAYEINDFRDGFYQIVNVQS-DAAAVQEFDRLAK 80

Query: 61 INGDILRHMIVK 72
          I+ DI+RH++VK
Sbjct: 81 ISDDIIRHIVVK 92
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 7215> which encodes the amino acid sequence <SEQ ID 7216>. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2720(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 66/74 (89%), Positives = 70/74 (94%)

Query: 1  MVARFDSILSDNGATVVESKDWEKRRLAYEIQDFTEGLYHIVNVEAEDAVALNEFDRLSK 60
          +VARFDSIL+DNGATVVESKDWEKRRLAYEI DF EGLYHIVN+EA DA ALNEFDRLSK
Sbjct: 22 LVARFDSILTDNGATVVESKDWEKRRLAYEINDFREGLYHIVNLEATDAAALNEFDRLSK 81

Query: 61 INGDILRHMIVKVD 74
          INGDILRHMIVK+D
Sbjct: 82 INGDILRHMIVKLD 95
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 2361**

A DNA sequence (GBSx2518) was identified in *S.agalactiae* <SEQ ID 7219> which encodes the amino acid sequence <SEQ ID 7220>. This protein is predicted to be surface protein Rib. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5289(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2362**

A DNA sequence (GASx1R) was identified in *S.pyogenes* <SEQ ID 7221> which encodes the amino acid sequence <SEQ ID 7222>. Analysis of this protein sequence reveals the following:

```
Possible site: 33

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2363**

A DNA sequence (GASx5R) was identified in *S.pyogenes* <SEQ ID 7223> which encodes the amino acid sequence <SEQ ID 7224>. Analysis of this protein sequence reveals the following:

```
Possible site: 20

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2364**

A DNA sequence (GASx11) was identified in *S.pyogenes* <SEQ ID 7225> which encodes the amino acid sequence <SEQ ID 7226>. Analysis of this protein sequence reveals the following:

```
Possible site: 22

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2614(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2365**

A DNA sequence (GASx17) was identified in *S.pyogenes* <SEQ ID 7227> which encodes the amino acid sequence <SEQ ID 7228>. Analysis of this protein sequence reveals the following:

```
Possible site: 30

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2849(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2366**

A DNA sequence (GASx18) was identified in *S.pyogenes* <SEQ ID 7229> which encodes the amino acid sequence <SEQ ID 7230>. Analysis of this protein sequence reveals the following:

```
Possible site: 30

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2099(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2367**

A DNA sequence (GASx34) was identified in *S.pyogenes* <SEQ ID 7231> which encodes the amino acid sequence <SEQ ID 7232>. Analysis of this protein sequence reveals the following:

```
Possible site: 54

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0801(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2368**

A DNA sequence (GASx38) was identified in *S.pyogenes* <SEQ ID 7233> which encodes the amino acid sequence <SEQ ID 7234>. Analysis of this protein sequence reveals the following:

```
Possible site: 18

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB12617 GB:Z99108 similar to protein-tyrosine phosphatase
          [Bacillus subtilis]
 Identities = 57/155 (36%), Positives = 88/155 (56%), Gaps = 12/155 (7%)

Query: 1   MKKVCFVCLGNICRSPMAEFVMKSIVS----SDVMMIESRATSDWEHGNPIHSGTQSILK 56
           M  V FVCLGNICRSPMAE + + + +       +  +S     W  GNP H GTQ IL+
Sbjct: 1   MISVLFVCLGNICRSPMAEAIFRDLAAKKGLEGKIKADSAGIGGWHIGNPPHEGTQEILR 60

Query: 57  TYQINYDITKCSKQITITDFNTFDYIIGMDSDNVKNLKEMSQHQWDSKIYLFRE------ 110
              I++D    ++Q++  D + FDYII MD++N+ +L+ M+  +  S I    +
Sbjct: 61  REGISFD-GMLARQVSEQDLDDFDYIIAMDAENIGSLRSMAGFKNTSHIKRLLDYVEDSD 119

Query: 111 -GGVPDPWYTNDFEETYQLVRKGCQDWLSRLMSKE 144
              VPDP+YT +FEE  QL++ GC+  L+ +  ++
Sbjct: 120 LADVPDPYYTGNFEEVCQLIKTGCEQLLASIQKEK 154
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2369**

A DNA sequence (GASx42R) was identified in *S.pyogenes* <SEQ ID 7235> which encodes the amino acid sequence <SEQ ID 7236>. Analysis of this protein sequence reveals the following:

```
Possible site: 14

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4753(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2370**

A DNA sequence (GASx47R) was identified in *S.pyogenes* <SEQ ID 7237> which encodes the amino acid sequence <SEQ ID 7238>. Analysis of this protein sequence reveals the following:

```
Possible site: 58

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2014(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2371**

A DNA sequence (GASx53R) was identified in *S.pyogenes* <SEQ ID 7239> which encodes the amino acid sequence <SEQ ID 7240>. Analysis of this protein sequence reveals the following:

```
Possible site: 45

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.11   Transmembrane   56 -  72 (  56 -  72)

----- Final Results -----
               bacterial membrane --- Certainty=0.1044(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2372**

A DNA sequence (GASx67R) was identified in *S.pyogenes* <SEQ ID 7241> which encodes the amino acid sequence <SEQ ID 7242>. Analysis of this protein sequence reveals the following:

```
Possible site: 39

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1610(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2373**

A DNA sequence (GASx75) was identified in *S.pyogenes* <SEQ ID 7243> which encodes the amino acid sequence <SEQ ID 7244>. Analysis of this protein sequence reveals the following:

```
Possible site: 31

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2803(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA41942 GB:X59250 ribosomal protein B [Lactococcus lactis]
 Identities = 37/38 (97%), Positives = 37/38 (97%)

Query: 1  MKVRPSVKPICEYCKVIRRNGRVMVICPTNPKHKQRQG 38
           MKVRPSVKPICEYCKVIRRNGRVMVICP NPKHKQRQG
Sbjct: 1  MKVRPSVKPICEYCKVIRRNGRVMVICPANPKHKQRQG 38
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2374**

A DNA sequence (GASx76) was identified in *S.pyogenes* <SEQ ID 7245> which encodes the amino acid sequence <SEQ ID 7246>. Analysis of this protein sequence reveals the following:

```
Possible site: 35

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
```

```
               bacterial cytoplasm --- Certainty=0.0824(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB06824 GB:L47971 ribosomal protein S13 [Bacillus subtilis]
 Identities = 86/121 (71%), Positives = 103/121 (85%)

Query: 1    MARIAGVDIPNDKRVVISLTYVYGIGLATSKKILAAAGISEDIRVKDLTSDQEDAIRREV 60
            MARIAGVDIP DKRVVISLTY++GIG  T++++L  AG+SED RV+DLT ++   IR  +
Sbjct: 1    MARIAGVDIPRDKRVVISLTYIFGIGRTTAQQVLKEAGVSEDTRVRDLTEEELGKIRDII 60

Query: 61   DAIKVEGDLRREVNMNIKRLMEIGSYRGIRHRRGLPVRGQNTKNNARTRKGKAVAIAGKKK 121
            D +KVEGDLRREV++NIKRL+EIGSYRGIRHRRGLPVRGQN+KNNARTRKG    +A KKK
Sbjct: 61   DKLKVEGDLRREVSLNIKRLIEIGSYRGIRHRRGLPVRGQNSKNNARTRKGPRRTVANKKK 121
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2375**

A DNA sequence (GASx81R) was identified in *S.pyogenes* <SEQ ID 7247> which encodes the amino acid sequence <SEQ ID 7248>. Analysis of this protein sequence reveals the following:

```
Possible site: 21

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.1842(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2376**

A DNA sequence (GASx82) was identified in *S.pyogenes* <SEQ ID 7249> which encodes the amino acid sequence <SEQ ID 7250>. Analysis of this protein sequence reveals the following:

```
Possible site: 59

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3613(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2377**

A DNA sequence (GASx83) was identified in *S.pyogenes* <SEQ ID 7251> which encodes the amino acid sequence <SEQ ID 7252>. Analysis of this protein sequence reveals the following:

```
Possible site: 51

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1141(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2378**

A DNA sequence (GASx85) was identified in *S.pyogenes* <SEQ ID 7253> which encodes the amino acid sequence <SEQ ID 7254>. Analysis of this protein sequence reveals the following:

```
Possible site: 16

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2280(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2379**

A DNA sequence (GASx89R) was identified in *S.pyogenes* <SEQ ID 7255> which encodes the amino acid sequence <SEQ ID 7256>. Analysis of this protein sequence reveals the following:

```
Possible site: 44

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3040(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2380**

A DNA sequence (GASx102) was identified in *S.pyogenes* <SEQ ID 7257> which encodes the amino acid sequence <SEQ ID 7258>. Analysis of this protein sequence reveals the following:

```
Possible site: 33

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-13.75    Transmembrane   21 -  37 (  12 -  41)

----- Final Results -----
               bacterial membrane --- Certainty=0.6498(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC45312 GB:U81957 ComYC [Streptococcus gordonii]
 Identities = 59/104 (56%), Positives = 85/104 (81%), Gaps = 1/104 (0%)

Query: 6    NNLRHKKLKGFTLLEMLLVILVISVLMLLFVPNLSKQKDRVTETGNAAVVKLVENQAELY 65
            N L+  ++K FTL+EML+V+L+ISVLMLLFVPNL+KQK+ V++TGNAAVVK+VE+QAELY
Sbjct: 2    NKLKKLRVKAFTLVEMLVVLLIISVLMLLFVPNLTKQKEAVSDTGNAAVVKVVESQAELY 61

Query: 66   EL-SQGSKPSLSQLKADGSITEKQEKAYQDYYDKHKNEKARLSN 108
            EL + G + +LS+L A G+I++KQ  +Y+ YY K+ +E   ++N
Sbjct: 62   ELKNTGDQATLSKLVAAGNISQKQADSYKAYYGKNNSETQAVAN 105
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2381**

A DNA sequence (GASx103) was identified in *S.pyogenes* <SEQ ID 7259> which encodes the amino acid sequence <SEQ ID 7260>. Analysis of this protein sequence reveals the following:

```
Possible site: 24

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC23740 GB:AF052207 competence protein [Streptococcus pneumoniae]
 Identities = 52/131 (39%), Positives = 76/131 (57%)

Query: 8    IKAFTLLETLLSLSVMSFIILGLSVPVTKSYQKVEEHLFFSHFEHLYRHQQKLAILQQKQ 67
            IKAFT+LE+LL L ++S + LGLS  V ++   VEE +FF  FE LYR  QK ++   Q++
Sbjct: 2    IKAFTMLESLLVLGLVSILALGLSGSVQSTFSAVEEQIFFMEFEELYRETQKRSVASQQK 61
```

```
Query: 68  RVLDISSTKIVTEGNSLTVPKSITVNHPYRLVIDQMGGNHSLAKIIFDMTDRRFKYQFYL 127
            L++     I       LTVPK I        +  D+ GGN SLAK+ F   +    +YQ YL
Sbjct: 62  TSLNLDGQMISNGSQKLTVPKGIQAPSGQSITFDRAGGNSSLAKVEFQTSKGAIRYQLYL 121

Query: 128 GSGNYQKTSQS 138
            G+G  ++  ++
Sbjct: 122 GNGKIKRIKET 132
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2382**

A DNA sequence (GASx104) was identified in *S.pyogenes* <SEQ ID 7261> which encodes the amino acid sequence <SEQ ID 7262>. Analysis of this protein sequence reveals the following:

```
Possible site: 23

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2383**

A DNA sequence (GASx109) was identified in *S.pyogenes* <SEQ ID 7265> which encodes the amino acid sequence <SEQ ID 7266>. Analysis of this protein sequence reveals the following:

```
Possible site: 45

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-10.51    Transmembrane    37 -   53 (   28 -   58)
   INTEGRAL    Likelihood = -3.56    Transmembrane    61 -   77 (   60 -   77)

----- Final Results -----
               bacterial membrane --- Certainty=0.5203(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2384**

A DNA sequence (GASx115R) was identified in *S.pyogenes* <SEQ ID 7267> which encodes the amino acid sequence <SEQ ID 7268>. Analysis of this protein sequence reveals the following:

```
Possible site: 18
```

```
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-11.09    Transmembrane    20 -   36 (   13 -   40)

----- Final Results -----
                 bacterial membrane --- Certainty=0.5437(Affirmative) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2385**

A DNA sequence (GASx124) was identified in *S.pyogenes* <SEQ ID 7269> which encodes the amino acid sequence <SEQ ID 7270>. Analysis of this protein sequence reveals the following:

```
Possible site: 52

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.17    Transmembrane    31 -   47 (   29 -   59)
   INTEGRAL    Likelihood = -5.63    Transmembrane   737 -  753 (  734 -  756)

----- Final Results -----
                 bacterial membrane --- Certainty=0.4270(Affirmative) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC97148 GB:U49397 Cpa [Streptococcus pyogenes]
 Identities = 401/737 (54%), Positives = 517/737 (69%), Gaps = 25/737 (3%)

Query: 25  SKNSKR--FTVTLVGVFLMIFALVTSMVGAKTVFGLVESSTPNAINPDSSSEYRWYGYES 82
            S N+KR   T+ L+ VFL   AL+  +  +   FG  E S PN     S  +Y WYGY+S
Sbjct: 11  SANNKRRQTTIGLLKVFLTFVALIGIVGFSIRAFGAEEQSVPN--RQSSIQDYPWYGYDS 68

Query: 83  YVRGHPYYKQFRVAHDLRVNLEGSRSYQVYCFNLKKAFPLGSDSSVKKWYKKHDGISTKF 142
            Y +G+P Y   +  H+L+VNLEGS+ YQ YCFNL K FP  SDS   +WYKK +G +  F
Sbjct: 69  YPKGYPDYSPLKTYHNLKVNLEGSKDYQAYCFNLTKHFPSKSDSVRSQWYKKLEGTNENF 128

Query: 143 EDYAMSPRITGDELNQKLRAVMYNGHPQNANGIMEGLEPLNAIRVTQEAVWYYSDNAPIS 202
               A  PRI   +L Q +  ++YNG+P N NGIM+G++PLNAI VTQ A+W Y+D+A I
Sbjct: 129 IKLADKPRIEDGQLQQNILRILYNGYPNNRNGIMKGIDPLNAILVTQNAIW-YTDSAQI- 186

Query: 203 NPDESFKRESESNLVSTSQLSLMRQALKQLIDPNLATKMPKQVPDDFQLSIFESEDKGDK 262
            NPDESFK E+ SN ++  QL LMR+ALK+LIDPNL +K   + P  ++L++FES D
Sbjct: 187 NPDESFKTEARSNGINDQQLGLMRKALKELIDPNLGSKYSNKTPSGYRLNVFESHD---- 242

Query: 263 YNKGYQNLLSGGLVPTKPPTPGDPPMPPNQPQTTSVLIRKYAIGDYSKLLEGATLQLTGD 322
              K +QNLLS   VP  PP PG+   PP + + TSV+IRKYA GD SKLLEGATL+L+
Sbjct: 243 --KPFQNLLSAEYVPDTPPKPGEE--PPAKTEKTSVIIRKYAEGD-SKLLEGATLKLSQI 297

Query: 323 NVNSFQARVFSSNDIGERIELSDGTYTLTELNSPAGYSIAEPITFKVEAGKVYTI-IDGK 381
              + FQ + F SN +GE +EL +GTYTLTE +SP GY IAEPI F+VE  KV+ +  DG
Sbjct: 298 EGSGFQEKDFQSNSLGETVELPNGTYTLTETSSPDGYKIAEPIKFRVENKKVFIVQKDGS 357

Query: 382 QIENPNKEIVEPYSVEAYNDFEEFSVLT-TQNYAKFYYAKNKNGSSQVVYCFNADLKSPP 440
            Q+ENPNKE+ EPYSVEAYNDF +  VL+    Y KFYYA NK+ SSQVVYCFNADL SPP
```

```
Sbjct: 358 QVENPNKEVAEPYSVEAYNDFMDEEVLSGFTPYGKFYYATNKDKSSQVVYCFNADLHSPP 417

Query: 441 DSEDGGKTMTPDFTT-GEVKYTHIAGRDLFKYTVKPRDTDPDTFLKHIKKVIEKGYREKG 499
           DS D G+T+ PD +T  EVKYTH AG DLFKY ++PRDT+P+ FLKHIKKVIEKGY++KG
Sbjct: 418 DSYDSGETINPDTSTMKEVKYTHTAGSDLFKYALRPRDTNPEDFLKHIKKVIEKGYKKKG 477

Query: 500 QAIEYSGLTETQLRAATQLAIYYFTDSAELDKDKL----KDYHGFGDMNDSTLAVAKILV 555
            +  Y+GLTETQ RAATQLAIYYFTDSA+L   K     K YHGF  M++ TLAV K L+
Sbjct: 478 DS--YNGLTETQFRAATQLAIYYFTDSADLKTLKTYNNGKGYHGFESMDEKTLAVTKELI 535

Query: 556 EYAQDSNPPQLTDLDFFIPNNNKYQSLIGTQWHPEDLVDIIRMEDKK-EVIPVTHNLTLR 614
            YAQ+ + PQLT+LDFF+PNN+K QSLIGT+ HP+DLVD+IRMEDKK EVIPVTH+LT++
Sbjct: 536 TYAQNGSAPQLTNLDFFVPNNSKDQSLIGTECHPDDLVDVIRMEDKKQEVIPVTHSLTVK 595

Query: 615 KTVTGLAGDRTKDFHFEIELKNNKQELLSQTVKTDKTNLEFKDGKATINLKHGESLTLQG 674
           KTV G  GD+TK F FE+ELK+   + +  T+KT+  +L  KDGK + NLKHG+++ ++G
Sbjct: 596 KTVVGELGDKTKGFQFELELKDKTGQPIVNTLKTNNQDLVAKDGKYSFNLKHGDTIRIEG 655

Query: 675 LPEGYSYLVKETDSEGYKVKVNSQEVANATVSKTGITSDETLAFENNKEPVVPTGVDQKI 734
           LP GYSY +KE +++ Y V V+++    A      IT D+ + FEN K+ V PTG+
Sbjct: 656 LPTGYSYTLKEAEAKDYIVTVDNKVSQEAQSVGKDITEDKKVTFENRKDLVPPTGLTTDG 715

Query: 735 NGYLALIVIAGISLGIW 751
             YL L+++  + L +W
Sbjct: 716 AIYLWLLLLVPLGLLVW 732
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2386**

A DNA sequence (GASx125R) was identified in *S.pyogenes* <SEQ ID 7271> which encodes the amino acid sequence <SEQ ID 7272>. Analysis of this protein sequence reveals the following:

```
Possible site: 40

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2604(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2387**

A DNA sequence (GASx126) was identified in *S.pyogenes* <SEQ ID 7273> which encodes the amino acid sequence <SEQ ID 7274>. Analysis of this protein sequence reveals the following:

```
Possible site: 14

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1537(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC97149 GB:U49397 LepA [Streptococcus pyogenes]
 Identities = 59/132 (44%), Positives = 84/132 (62%), Gaps = 5/132 (3%)

Query: 1   MIIKRNDMAPSVKAGDAILFYRLSQTYKVEEAVVYEDSKTSITKVGRIIAQAGDEVDLTE 60
           MII  NDM+P++ AGD +L+YRL+    + + VVYE   T   KVGRI AQAGDEV+ T+
Sbjct: 42  MIINTNDMSPALSAGDGVLYYRLADRSHINDVVVYEVDNT--LKVGRIAAQAGDEVNFTQ 99

Query: 61  QGELKINGHIQNEG---LTFIKSREANYPYRIADNSYLILNDYYSQESENYLQDAIAKDA 117
           +G L INGH   +    LT+  S   N+PY++   +Y ILNDY  +  ++    A+  +
Sbjct: 100 EGGLLINGHPPEKEVPYLTYPHSSGPNFPYKVPTGTYFILNDYREERLDSRYYGALPINQ 159

Query: 118 IKGTINTLIRLR 129
           IKG I+TL+R+R
Sbjct: 160 IKGKISTLLRVR 171
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2388**

A DNA sequence (GASx127) was identified in *S.pyogenes* <SEQ ID 7275> which encodes the amino acid sequence <SEQ ID 7276>. Analysis of this protein sequence reveals the following:

```
Possible site: 17

>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -3.93    Transmembrane  312 - 328 ( 311 - 337)

----- Final Results -----
               bacterial membrane --- Certainty=0.2572(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC97152 GB:U49397 unknown [Streptococcus pyogenes]
 Identities = 125/355 (35%), Positives = 191/355 (53%), Gaps = 26/355 (7%)

Query: 1   MKLRHLLLTGAALTSFA-----ATTVHGET--VVNGAKLTVTKNL-DLVNSNALIPNTDF 52
           MK   LLL  A L +       +  +  ET  V++G+ L V K      + N L+P  D+
Sbjct: 1   MKKNKLLLATAILATALGMASMSQNIKAETAGVIDGSTLVVKKTFPSYTDDNVLMPKADY 60

Query: 53  TFKIEPDTTVN---EDGNKFK-GVALNTPMTK-VTYTNSDKGGSNTKTAEFDFSEVTFEK 107
           +FK+E D       +DG   K GV      TK + Y+NSDK   +  K+  F+F+ V F
Sbjct: 61  SFKVEADDNAKGKTKDGLDIKPGVIDGLENTKTIRYSNSDKITAKEKSVNFEFANVKFPG 120

Query: 108 PGVYYYKVTEEKIDKVPGVSYDTTSYTVQVHVLWNEEQQKPVATYIVGYKEGS--KVPIQ 165
            GVY Y V E   +K  G++YD+  +TV V+V+ N+E       YIV  + G   K P+
Sbjct: 121 VGVYRYTVAEVNGNKA-GITYDSQQWTVDVYVV-NKEGGGFEVKYIVSTEVGQSEKKPVL 178

Query: 166 FKNSLDSTTLTVKKKVSGTGGDRSKDFNFGLTLKANQYYKASEKVMIEKTTKGGQAPVQT 225
           FKNS D+T+L ++K+V+G  G+  + F+F L L  N+ +   EK  +    +GG+
Sbjct: 179 FKNSFDTTSLKIEKQVTGNTGEHQRLFSFTLLLTPNECF---EKGQVVNILQGGETK--- 232

Query: 226 EASIDQLYHFTLKDGESIKVTNLPVGVDYVVTEDDYKSEKYTTNVEVSPQDGAVKNIAGN 285
           +  I + Y FTLKD  S+ ++ LPVG++Y +TE+D   + Y T+  +   + +     G
Sbjct: 233 KVVIGEEYSFTLKDKGSVTLSQLPVGIEYKLTEEDVTKDGYKTSATLKDGEQSSTYELGK 292

Query: 286 STEQETSTDKDMTITFTNKKDFEVPTGVAMTVAPYIALGIVAVGGALYFVKKKNA 340
             + + S D+   I  TNK+D +VPTGV  T+AP+  L IVA+GG +Y  K+K A
```

```
Sbjct: 293 DHKTDKSADE---IVVTNKRDTQVPTGVVGTLAPFAVLSIVAIGGVIYITKRKKA 344
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

## Example 2389

A DNA sequence (GASx128) was identified in *S.pyogenes* <SEQ ID 7277> which encodes the amino acid sequence <SEQ ID 7278>. Analysis of this protein sequence reveals the following:

```
Possible site: 44

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC97152 GB:U49397 unknown [Streptococcus pyogenes]
 Identities = 115/240 (47%), Positives = 178/240 (73%), Gaps = 3/240 (1%)

Query: 1   MIVRLIKLLDKLINVIVLCFFFLCLLIAALGIYDALTVYQGANATNYQQYKKKGVQ--FD 58
           M++ ++++++K I+ ++L F  + L +A  G++D+  +YQ A+A+N++++K    Q  F+
Sbjct: 351 MMMTIVQVINKAIDTLILIFCLVVLFLAGFGLWDSYHLYQQADASNFKKFKTAQQQPKFE 410

Query: 59  DLLAINSDVMAWLTVKGTHIDYPIVQGENNLEYINKSVEGEYSLSGSVFLDYRNKVTFED 118
           DLLA+N DV+ WL + GTHIDYP+VQG+ NLEYINK+V+G  ++SGS+FLD RN    F D
Sbjct: 411 DLLALNEDVIGWLNIPGTHIDYPLVQGKTNLEYINKAVDGSVAMSGSLFLDTRNHNDFTD 470

Query: 119 KYSLIYAHHMAGNVMFGELPNFRKKSFFNKHKEFSIETKTKQKLKINIFACIQTDAFDSL 178
            YSLIY HHMAGN MFGE+P F KK+FFNKH +  IETK ++KL + IFAC++TDAFD L
Sbjct: 471 DYSLIYGHHMAGNAMFGEIPKFLKKNFFNKHNKAIIETKERKKLTVTIFACLKTDAFDQL 530

Query: 179 LFNPIDV-DISSKNEFLNHIKQKSVQYREILTTNESRFVALSTCEDMTTDGRIIVIGQIE 237
           +FNP  + +   + + +++I ++S Q++ +    + ++FVA STCE+ +TD R+IV+G I+
Sbjct: 531 VFNPNAITNQDQQRQLVDYISKRSKQFKPVKLKHHTKFVAFSTCENFSTDNRVIVVGTIQ 590
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

## Example 2390

A DNA sequence (GASx129) was identified in *S.pyogenes* <SEQ ID 7279> which encodes the amino acid sequence <SEQ ID 7280>. Analysis of this protein sequence reveals the following:

```
Possible site: 26

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -6.05    Transmembrane    5 -  21 (   4 -  22)
   INTEGRAL    Likelihood = -5.04    Transmembrane  191 - 207 ( 186 - 209)

----- Final Results -----
               bacterial membrane --- Certainty=0.3421(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

LPXTG motif: 181-186
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC97151 GB:U49397 unknown [Streptococcus pyogenes]
 Identities = 64/213 (30%), Positives = 106/213 (49%), Gaps = 20/213 (9%)

Query: 1   MKKSILRILAIGYLLMSFCLLDSVEAENLTASINIEVINQVDVATNKQSSDIDETFMFVI 60
            M+K    + ++  +L       +V A++ T   +I V N ++ A +        F   +
Sbjct: 1   MRKYWKMLFSVVMMLTMLAFNQTVLAKDSTVQTSISVENVLERAGDSTP------FSIAL 54

Query: 61  EALDKESPLPNSVTTSVKGNGKTSFEQLTFSEVGQYHYKIHQLLGKNSQYHYDETVYEVV 120
            E++D    +      ++ G+GK SF  L F+ VGQY Y+++Q   +N  Y  D TV++V+
Sbjct: 55  ESIDAMKTIEE---ITIAGSGKASFSPLNFTTVGQYTYRVYQKPSQNKDYQADTTVFDVL 111

Query: 121 IYVLYNEQSGALETNLVSNKLGETEKSELIFKQEYSEKTPEPHQPDTTEKEKPQKKRNGI 180
            +YV Y+E  G L   ++S + G+ EKS + FK +   K   P QPD  +
Sbjct: 112 VYVTYDE-DGTLVAKVISRRAGDEEKSAITFKPKRLVKPIPPRQPDIPKTP--------- 161

Query: 181 LPSTGEMVSYVSALGIVLVATITLYSIYKKLKT 213
            LP  GE+ S +  L IVL+  + L  + KKLK+
Sbjct: 162 LPLAGEVKSLLGILSIVLLGLLVLLYV-KKLKS 193
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2391**

A DNA sequence (GASx130R) was identified in *S.pyogenes* <SEQ ID 7281> which encodes the amino acid sequence <SEQ ID 7282>. Analysis of this protein sequence reveals the following:

```
Possible site: 57

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1614(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB54046 GB:AJ245436 hypothetical protein, 57.8 kD [Pseudomonas
           putida]
 Identities = 128/388 (32%), Positives = 204/388 (51%), Gaps = 21/388 (5%)

Query: 4   IGSVVQRQELVFIPAQLKRINHVQHAYKCQTCSDNSLSDKIIKAPVPKAPLAHSLGSASI 63
            IG  V  Q L  +P Q++ I HV+  Y C+ C    ++     A  P   +  S+ S S+
Sbjct: 126 IGEEVSEQ-LEIVPMQIRVIKHVRKVYGCRDCESAPVT-----ADKPAQMIEKSMASPSV 179

Query: 64  IAHTVHQKFTLKVPNYRQEEDWNKLGLSISRKEIANWHIKSSQYYFEPLYDLLRDILLSQ 123
            +A  +  K+   +P +R E+   + G+ I R+ +A W I+ S++ F+PL +L+R+ LL+
Sbjct: 180 LAMLLTTKYVDGLPLHRFEKVLGRHGIDIPRQTLARWVIQCSEH-FQPLLNLMRESLLNS 238

Query: 124 EVIHADETSYRVLESD----TQLTYYWTFLSGKHEKKGITLYHHDKRRSGLVTQEVLGDY 179
             +IH DET  +VL+      +  ++ W    G  ++  + L+ +   R+  V   +L  Y
Sbjct: 239 RIIHCDETRVQVLKEPGREPSSQSWMWVQTGGPPDRP-VILFDYATSRAQEVPVRLLDGY 297

Query: 180 SGYVHCDMHGAYRQL---EHAKLVGCWAHVRRKFFEATPKQAD-KTSLGRKGLVYCDKLF 235
             GYV  D +  Y  L   +  + +GCWAH RRKF EA   Q   KT      L   +KL+
Sbjct: 298 RGYVMTDDYAGYNALAAQDGLERLGCWAHARRKFVEAQKVQPKGKTGRADIALNLINKLY 357

Query: 236 ALEAEWCELPPQERLVKRKEILTPLMTTFFDWCR--EQVVLSGSKLGLAIAYSLKHERTF 293
             +E +  +   ++R V R E   PL+T   +W    +  V + + LG AI Y   +
```

```
Sbjct: 358 GVERDLKDSDDEDRKVARMERSLPLLTQLKNWVEKTQPQVTTQNALGKAIGYLASNWSKL 417

Query: 294 RTVLEDGHIVLSNNMAERAIKSLVMGRKNWLFSQSFEGAKAAAIIMSLLETAKRHGLNSE 353
              +E G++ + NN AERAI+  V+GRKNWLFS + +GA A+A + SL+ETAK +G
Sbjct: 418 ERYVEHGYLPMDNNAAERAIRPFVIGRKNWLFSDTPKGATASAQLYSLVETAKANGQEPY 477

Query: 354 KYISYLLDRLPNEETLAKREVLEAYLPW 381
             ++ + L+RLP    ++    E  EA LPW
Sbjct: 478 AWLRHALERLPQACSV---EDYEALLPW 502
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2392**

A DNA sequence (GASx131R) was identified in *S.pyogenes* <SEQ ID 7283> which encodes the amino acid sequence <SEQ ID 7284>. Analysis of this protein sequence reveals the following:

```
Possible site: 37

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4465(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2393**

A DNA sequence (GASx132R) was identified in *S.pyogenes* <SEQ ID 7285> which encodes the amino acid sequence <SEQ ID 7286>. Analysis of this protein sequence reveals the following:

```
Possible site: 46

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1529(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAA84885 GB:AB024946 orf50 [Escherichia coli]
 Identities = 37/91 (40%), Positives = 53/91 (57%)

Query: 10  QVYLVCGKTDMRQGIDSLAYLVKSQHELDLFSGAVYLFCGGRRDRFKALYWDGQGFWLLY 69
            +++LV G TDMR G + LA  V++  + D FSG +++F G R D+ K L+ D  G  L
Sbjct: 9   RIWLVAGITDMRNGFNGLASKVQNVLKDDPFSGHLFIFRGRRGDQIKVLWADSDGLCLFT 68

Query: 70  KRFENGKLAWPRNRDEVKCLTAVQVDWLMKG 100
            KR E G+  WP  RD     LT  Q+  L++G
Sbjct: 69  KRLERGRFVWPVTRDGKVHLTPAQLSMLLEG 99
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2394

A DNA sequence (GASx133R) was identified in *S.pyogenes* <SEQ ID 7287> which encodes the amino acid sequence <SEQ ID 7288>. Analysis of this protein sequence reveals the following:

```
Possible site: 18

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1979(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2395

A DNA sequence (GASx135R) was identified in *S.pyogenes* <SEQ ID 7289> which encodes the amino acid sequence <SEQ ID 7290>. Analysis of this protein sequence reveals the following:

```
Possible site: 20

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2396

A DNA sequence (GASx136) was identified in *S.pyogenes* <SEQ ID 7291> which encodes the amino acid sequence <SEQ ID 7292>. Analysis of this protein sequence reveals the following:

```
Possible site: 54

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-11.73   Transmembrane  222 - 238 ( 212 - 242)
   INTEGRAL    Likelihood =-10.88   Transmembrane   37 -  53 (  32 -  57)
   INTEGRAL    Likelihood = -9.87   Transmembrane  462 - 478 ( 456 - 478)
   INTEGRAL    Likelihood = -4.25   Transmembrane  119 - 135 ( 117 - 137)
   INTEGRAL    Likelihood = -2.60   Transmembrane  308 - 324 ( 306 - 324)
   INTEGRAL    Likelihood = -1.28   Transmembrane  164 - 180 ( 164 - 180)
   INTEGRAL    Likelihood = -0.06   Transmembrane  137 - 153 ( 137 - 153)
   INTEGRAL    Likelihood = -0.06   Transmembrane  343 - 359 ( 343 - 359)
```

```
----- Final Results -----
               bacterial membrane --- Certainty=0.5692(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB04077 GB:AP001508 short-chain fatty acids transporter
          [Bacillus halodurans]
 Identities = 158/465 (33%), Positives = 248/465 (52%), Gaps = 41/465 (8%)

Query: 15  IKTKKRFMDRYIDGFMKWMPESLFICFILTFLVVTMSVLMTDSPFIGTEKTGGIIYGWVN 74
            I    R M RY+       P+      +LTFLV  +S++ T+S    T  T  I+  W
Sbjct: 5   ISLSNRLMQRYL-------PDPFLFVVLLTFLVFALSLIFTES----TPLT--IVQYWGE 51

Query: 75  GFWGLLSFAMQMTILLATGNAVASSPPAHKMFKSLAKLPQTRTQIFIFSIVVGSIFGFLH 134
            GFWGLLSF+MQM ++L TG+ +ASSP   K   +LA LP +  Q  +   VV  +  F++
Sbjct: 52  GFWGLLSFSMQMVLVLVTGHVLASSPLFKKGLGALAGLPASPGQAILLVTVVSLVASFIN 111

Query: 135 WGLGMMVAIVFGKELLVQARQKGIKVHTPLFVATLFFTFLPATSGLSGAAVLYSATPDYL 194
            WG G+++  +F KEL     +K   V   L +A+ +  F+    GLSG+  L  ATPD+
Sbjct: 112 WGFGLVIGALFAKELA----KKVDNVDYRLLIASAYSGFMIWHGGLSGSVPLTIATPDHF 167

Query: 195 RNSVADAYKQVVPESVPLTESVL---NLPFISLLVVCMLVPLCFALLAHPKDETKIME-- 249
               +           +P +E++    NL  +  L +   +PL   L+   K +T  ++
Sbjct: 168 AQDMIGV--------IPTSETIFAPYNLAIVFALFIA--IPLANRLMMPGKSDTVTVDRS 217

Query: 250 -LDDEIYHHSLDTASHVVIARNTPAEKMNASRLVMYLVGGAIVSYSLYHFSVVGLSGLDL 308
             LDD      L  AS + +   TP++++  SR++  LVG   + +  Y+F+  G   L+L
Sbjct: 218 LLDDG---RDLQAAS-LELEAMTPSDRLENSRMISLLVGVLGLVFLGYYFATNGFE-LNL 272

Query: 309 NCFNFLFLGLGLLLCGQQGPEYYGSLFKDGVMSSWGLVLQFPFYAGIFGIIQSTGLGLEI 368
           +  N LFL LG+L  G   P+ +       V  + G+++QFPFYAG+ GI+ S+GL   +
Sbjct: 273 DIVNSLFLFLGILFHGT--PKLFLKAVTSAVKGASGIIIQFPFYAGLMGIMVSSGLATVM 330

Query: 369 SHFFVAISNGTTWPVFAYLYSALLNIAVPSGGSKFVIEAPYIVPATIEVGNDLGKILQAY 428
           S  FV+ SN  T+P+F +L + ++N+ VPSGG ++ ++AP ++ A   +G    K   A
Sbjct: 331 SEAFVSFSNEVTFPLFVFLSAGIVNVFVPSGGGQWAVQAPVVLEAAQSLGVPAAKAAMAV 390

Query: 429 QLGDATTNLIVPFWALSYLSNFKLKFNQIVAYTIPCVLVVTGIAI 473
             GDA TN+I PFWAL  L+   LK   I+ + +  +LVV+G+ I
Sbjct: 391 AWGDAWTNMIQPFWALPALAIAGLKAKDIMGFCV-MILVVSGVVI 434
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2397**

A DNA sequence (GASx137R) was identified in *S.pyogenes* <SEQ ID 7293> which encodes the amino acid sequence <SEQ ID 7294>. Analysis of this protein sequence reveals the following:

```
Possible site: 58

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2591(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC22434 GB:U32761 transcriptional regulator [Haemophilus influenzae Rd]
 Identities = 37/107 (34%), Positives = 56/107 (51%), Gaps = 1/107 (0%)

Query: 21  LHRQNLVTFDKTFMINHQLTTLFEEANSLPVVKCYSASWDFLLNCTRYS-SYLTILPRPI 79
            LH+Q +  FD+TFMI+H L   FE  N  P +   S+ WDFLL+ +  +   LTILP P+
Sbjct: 205 LHQQKMAIFDQTFMIHHHLKEAFERNNCYPDIVLDSSCWDFLLSAVKTNKELLTILPLPM 264

Query: 80  THFAHMDGLVEVQLTEHPKWEVVLASLKHNKTSHLKHYIKHTILDYF 126
                H    +  ++     W+V L   +    +HL+ YI   +L+ F
Sbjct: 265 AELYHSKEFLCRKIESPVPWKVTLCRQRKTVYTHLEEYIFDKLLEAF 311
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2398**

A DNA sequence (GASx140) was identified in *S.pyogenes* <SEQ ID 7295> which encodes the amino acid sequence <SEQ ID 7296>. Analysis of this protein sequence reveals the following:

```
Possible site: 50

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3351(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
!GB:U32761 acetate CoA-transferase, alpha subunit [H...    215  4e-55
 Identities = 105/213 (49%), Positives = 146/213 (68%)

Query: 22  ENKRIAIAEAISHIKDGDTIMVGGFMANGTPEALIDALVDKGTKDLTLICNDAGFVDRGV 81
            + K + + +A    +DG TIMVGGFM  GTP  L++AL++ G +DLTLI ND  FVD G+
Sbjct: 2   KTKLMTLQDATGFFRDGMTIMVGGFMGIGTPSRLVEALLESGVRDLTLIANDTAFVDTGI 61

Query: 82  GKMVANHQFKTIYATHIGLNKEAGRQMTAGETTIELIPQGTFAEKIRIGAYGIGGFYTPT 141
            G ++ N + + + A+HIG N E GR+M +GE  + L+PQGT  E+IR G  G+GGF TPT
Sbjct: 62  GPLIVNGRVRKVIASHIGTNPETGRRMISGEMDVVLVPQGTLIEQIRCGGAGLGGFLTPT 121

Query: 142 GVGTLVAEGKETKTIKGKTYLLEYPFEADVALIFANQADEMGNLQYSGSENNFNQLMAAC 201
            GVGT+V EGK+T T+ GKT+LLE P  AD+ALI A++ D +GNL Y  S  NFN L+A
Sbjct: 122 GVGTVVEEGKQTLTLDGKTWLLERPLRADLALIRAHRCDTLGNLTYQLSARNFNPLIALA 181

Query: 202 AKTTIVQAREIVPVGTIQPECVHTPHIFVDYIV 234
            A  T+V+  E+V  G +QP+ + TP   +D+I+
Sbjct: 182 ADITLVEPDELVETGELQPDHIVTPGAVIDHII 214
            subunit (EC 2.8.3.-). [Escherichia coli]
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2399**

A DNA sequence (GASx141) was identified in *S.pyogenes* <SEQ ID 7297> which encodes the amino acid sequence <SEQ ID 7298>. Analysis of this protein sequence reveals the following:

```
Possible site: 41

>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4941(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF12248 GB:AE001862 CoA transferase, subunit B [Deinococcus radiodurans]
 Identities = 114/203 (56%), Positives = 158/203 (77%), Gaps = 3/203 (1%)

Query: 11  QNRIAKRVAKELEDGTLVNLGIGLPTKVANFVPEEMTVYFQSENGFIGLGP--KSDDPNS 68
            ++ +A R A+EL+DG  VNLGIGLPT VAN +P  M+V+ QSENG +G+GP    D+ +
Sbjct: 5   RDEMAARAAQELQDGYYVNLGIGLPTLVANHIPAGMSVWLQSENGLLGIGPFPTEDEVDP 64

Query: 69  TIVNAGGQPVTVYPGAAFFNSADSFGIIRGGHVDLTVLGALEIAENGDIANYLIPGKMVP 128
             ++NAG Q VT  PGA+FF+SADSF +IRGGHV+L +LGA++++E GD+AN++IPGKMV
Sbjct: 65  DLINAGKQTVTALPGASFFSSADSFAMIRGGHVNLAILGAMQVSETGDLANWMIPGKMVK 124

Query: 129 GMGGAMDLLVGAKKVIVAMEHTNKG-KHKLLKECTLPLTAKGVVDLIITEMGVFKVTPDG 187
            GMGGAMDL+ G ++V+V MEH  KG  HK+L+ECTLPLT +GVVD IIT++GV  VTP G
Sbjct: 125 GMGGAMDLVAGVQRVVVLMEHVAKGDAHKILRECTLPLTGQGVVDRIITDLGVLDVTPQG 184

Query: 188 IQVIEISEGFTFDEVQAATGVPL 210
            ++++E++ G T DE++  TG  +
Sbjct: 185 LKLVELAPGVTLDELRQKTGADI 207
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2400**

A DNA sequence (GASx144) was identified in *S.pyogenes* <SEQ ID 7299> which encodes the amino acid sequence <SEQ ID 7300>. Analysis of this protein sequence reveals the following:

```
Possible site: 39

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3227(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAA29948 GB:AP000003 137aa long hypothetical protein [Pyrococcus
           horikoshii]
 Identities = 49/113 (43%), Positives = 71/113 (62%), Gaps = 1/113 (0%)

Query: 5   PEPMGPYSTYTIEGHFLYTAGQLPLNPVTGQLSDG-FEAQCRQVFVNLQSILAEQKLDLN 63
            P+P+GPYS     G+FL+ AGQ+P++P TG++  G  + Q RQV  N+++IL      LN
Sbjct: 22  PKPIGPYSQAIKAGNFLFIAGQIPIDPKTGEIVKGDIKDQTRQVLENIKAILEAAGYSLN 81

Query: 64  HIYKLNVYLTDVTNVEILNHVMTDLFEEPYPVRTAVQVSALPLQALIEVEAVA 116
             + K+ VYL D+ +   +N V  + F E  P R AV+VS LP   LIE+EA+A
Sbjct: 82  DVIKVTVYLKDMNDFAKMNEVYAEYFGESKPARVAVEVSRLPKDVLIEIEAIA 134
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2401**

A DNA sequence (GASx146) was identified in *S.pyogenes* <SEQ ID 7301> which encodes the amino acid sequence <SEQ ID 7302>. Analysis of this protein sequence reveals the following:

```
Possible site: 16

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.1238(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2402**

A DNA sequence (GASx147) was identified in *S.pyogenes* <SEQ ID 7303> which encodes the amino acid sequence <SEQ ID 7304>. Analysis of this protein sequence reveals the following:

```
Possible site: 30

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-11.46   Transmembrane  456 - 472 ( 452 - 481)
   INTEGRAL    Likelihood = -8.17   Transmembrane  603 - 619 ( 595 - 623)
   INTEGRAL    Likelihood = -6.85   Transmembrane  495 - 511 ( 491 - 518)
   INTEGRAL    Likelihood = -5.31   Transmembrane  420 - 436 ( 418 - 443)
   INTEGRAL    Likelihood = -4.99   Transmembrane  396 - 412 ( 392 - 413)
   INTEGRAL    Likelihood = -1.59   Transmembrane  522 - 538 ( 522 - 538)
   INTEGRAL    Likelihood = -0.64   Transmembrane  577 - 593 ( 577 - 593)
   INTEGRAL    Likelihood = -0.43   Transmembrane  377 - 393 ( 377 - 393)

----- Final Results -----
                bacterial membrane --- Certainty=0.5585(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAA04270 GB:D17462 Na+ -ATPase subunit I [Enterococcus hirae]
 Identities = 232/681 (34%), Positives = 370/681 (54%), Gaps = 40/681 (5%)

Query: 1   MAISQMKKLAMVFEKDYLDLVLKTLQQSQLVEVRDMKQLKH---WQDAFNKGNVKLPQIV 57
           MA+++M+K+ ++ +K   +++L+ +Q    VE+RD+ Q      W + F       P+++
Sbjct: 1   MAVTKMEKVTLISDKKNREILLQAVQGLHAVEIRDLFQESENNQWVETF----FPEPEMI 56

Query: 58  QYDLTHQKPLLDDEALQYLLQSQQELENGLASLSAFLPPIGKLTALRQ--KTPSLSFKQF 115
             D    K       L Y L         + +   F+   G+ +  +Q  K   LS
Sbjct: 57  DKDKELAK-------LSYKLTD-------IRTAIQFIEHHGEKSQKKQHLKRRELSLDTL 102

Query: 116 EERHRQQAAQTALKMMSQKIERLEQLQSKIDQLTEYCQELEKWRSLTVLPQDLAQFHFLS 175
           E+ + ++A    L+ +    E+ EQL  +  QL +    L  W++L + P+
Sbjct: 103 EKNYSEEAFSKKLEEVLLLKEQWEQLVDERQQLEDQENWLLNWQNLDLAPKAFDS-QMTK 161

Query: 176 ARVGTIPSTANNHFYHQLKQHKGLFIEEVYH----TEFEYGLVLFWQAQDTIHLQKYQFK 231
             +GT+ +     F  ++ +    ++EE+      T F Y ++   +++      +Y F
Sbjct: 162 LVIGTVNAKNAESFKAEVAEINEAYLEEINSSPTTTYFAYIVLRADESRMEEIASRYGFV 221
```

```
Query: 232 PLLYKEQLLPSEQLRINKELLTNWLAEKDSLLKELRQSQKILAQLQVEIDYVLSQYQRQQ 291
              Y  +  P +QL   K+ L     ++  L   +      +   +   +  L++ +R+
Sbjct: 222 KEDYLYEGTPQQQLVAAKQSLQEIKDQQKKLSSAIGACSGYIKDFEWTEEIFLARSEREA 281

Query: 292 TKKQLLGTRHLIALEGWIEADSVNQLKGLMTKTLGDMFYLDSYDVTPDDW--EDVPIKLR 349
            K +++ T +LI ++GW++ +   +L  ++   L       ++D   D+   E+VP KL+
Sbjct: 282 IKDRIIHTPYLILIQGWVDHEEKQELIHMLQNILASEEVYLTFDEPTDNEIAEEVPTKLK 341

Query: 350 NHRYIAPFELVTEMYALPKYQEKDPTPFLAPLYLTFFGMMVADLGYGLLLYAVTLAALVF 409
           NH  +APFE++TEMY+LPKY+E DPTP++ P YL FFGMMVAD+GYGLL++
Sbjct: 342 NHPIVAPFEMLTEMYSLPKYEEVDPTPWMMPFYLVFFGMMVADIGYGLLMFLGAFLLQKL 401

Query: 410 FNLQKTSKRLVTFFNILAISVAIWGLIYGSFFG---------FDLPVALLSTKTDVITIL 460
             L +  +R   FF ILAI   IWG IY SFFG           LP  +LST  DV TIL
Sbjct: 402 VVLPRGMQRFAKFFEILAIPSIIWGFIYSSFFGAALPKEIFGIHLPFPILSTTDDVNTIL 461

Query: 461 VVSLLFGFVTLIFGLLLGAWQQVRMKAYATAYTSSLAWTFILLGLLLFILGKNVSGLAYL 520
           ++S++FG + ++ GL + A + ++ KAY  A     AW +ILLG++L +LG
Sbjct: 462 ILSVIFGLIQILVGLFIAAKEHIKRKAYVDAVNDGFAWQWILLGIILILLGTMTLKNNAF 521

Query: 521 SVIGKWLALGNAFGILVVSLLKSKSLL-GLGSGLYNLYGISSYLSDLVSFTRLMALGLSG 579
             +G  LA+ +A  IL++ + +S S   G+  G YNLYG++ Y+ DLVS+TRLMALG+SG
Sbjct: 522 VYLGGALAVLSAVCILIIPVFQSSSKAKGIAKGAYNLYGLTGYIGDLVSYTRLMALGISG 581

Query: 580 ASIGAAFNMIVGIFPPVTRFTVGIFIFILLHAINIFLSMLSGYVHGARLIFVEFFGKFYE 639
            SI AAFNM+V   PP  RF+VGI + I+L A+N+FL++LS YVHGARL +VEFFGKFY
Sbjct: 582 GSIAAAFNMLVAFMPPAARFSVGILLIIVLQALNMFLTLLSAYVHGARLQYVEFFGKFYT 641

Query: 640 GGGKAFNPLKLADNYVNVNEE 660
           GGG++F PLK  + YVN+N +
Sbjct: 642 GGGRSFKPLKTVEKYVNINHK 662
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2403**

A DNA sequence (GASx148) was identified in *S.pyogenes* <SEQ ID 7305> which encodes the amino acid sequence <SEQ ID 7306>. Analysis of this protein sequence reveals the following:

```
Possible site: 40

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -7.80   Transmembrane   28 -  44 (  21 -  51)
   INTEGRAL    Likelihood = -6.85   Transmembrane  148 - 164 ( 146 - 170)
   INTEGRAL    Likelihood = -2.81   Transmembrane  105 - 121 ( 105 - 123)

----- Final Results -----
               bacterial membrane --- Certainty=0.4121(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAA03841 GB:D16334 Na+-ATPase K subunit [Enterococcus hirae]
 Identities = 85/150 (56%), Positives = 107/150 (70%)

Query: 20  HYFTAHGGVFFAALGIVLAVALSGMGSAYGVGKGGQAAAALLKEEPEKFTSALILQLLPG 79
           +  T +GG+ FA L + A   SG+GSA GVG  G+AAAAL   +PEKF  ALILQLLPG
Sbjct: 4   YLITQNGGMVFAVLAMATATIFSGIGSAKGVGMTGEAAAALTTSQPEKFGQALILQLLPG 63

Query: 80  SQGIYGFAIGILIWMKLTPELSVNQGLAYFLVSLPIAIVGYFSAKHQGNVSVAGMQILAK 139
           +QG+YGF I  LI++ L  ++SV QGL +   SLPIA  G FS   QG V+ AG+QILAK
```

```
Sbjct: 64  TQGLYGFVIAFLIFINLGSDMSVVQGLNFLGASLPIAFTGLFSGIAQGKVAAAGIQILAK 123

Query: 140 RPKDFMKGVILAAMVETYAILAFVVSFILL 169
           +P+   KG+I AAMVETYAIL FV+SF+L+
Sbjct: 124 KPEHATKGIIFAAMVETYAILGFVISFLLV 153
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2404**

A DNA sequence (GASx149) was identified in *S.pyogenes* <SEQ ID 7307> which encodes the amino acid sequence <SEQ ID 7308>. Analysis of this protein sequence reveals the following:

```
Possible site: 55

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4510(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAA04272 GB:D17462 Na+ -ATPase subunit E [Enterococcus hirae]
 Identities = 43/193 (22%), Positives = 95/193 (48%), Gaps = 2/193 (1%)

Query: 1   VNDITQLRQNVLEKAHQEGQQCLKIATDSLDTDFKERQQQGLHDLKAKRQKELKALEQQF 60
           V+ I ++   + E A  E     ++    +D  F+ ++ Q   D + ++  +L+ +E+ +
Sbjct: 3   VDAIDKIITQINETAQLERASFEEMKRKEIDQKFEVKKWQIEADFQKEKASKLEEIERSY 62

Query: 61  QVAQQQLKNQERQALLALKQDSIKELFEASLEKMTNFSKEEELAFLKQVLSKYP-EQPLQ 119
           +  + + K Q +Q +L  KQ+ ++ LF  +  ++ N  KEE+LA +KQ++   P     +
Sbjct: 63  RQLRNKQKMQVKQEILNAKQEVLQRLFTEATLQLENEPKEEQLALMKQMIQTLPINGTAR 122

Query: 120 VTFGEKTGQKFSSYDCAELRLAFPQLSYNQELIPQ-EAGFLVSLDQVDDNYLYRYLLESV 178
           +  GEK+    +    AE     P     ++   + +AG ++    +  N+L+ +L++ +
Sbjct: 123 LIPGEKSADILTPAVIAEWNEELPFELIREDFTEKAQAGLIIDDAGIQYNFLFSHLIKEI 182

Query: 179 LKEESSRIIDMLF 191
            +  S+ I   LF
Sbjct: 183 QETMSAEIAKELF 195
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2405**

A DNA sequence (GASx150) was identified in *S.pyogenes* <SEQ ID 7309> which encodes the amino acid sequence <SEQ ID 7310>. Analysis of this protein sequence reveals the following:

```
Possible site: 30

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3095(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAA04273 GB:D17462 Na+ -ATPase subunit C [Enterococcus hirae]
 Identities = 94/326 (28%), Positives = 167/326 (50%), Gaps = 5/326 (1%)

Query: 6   ELNTTISVKEKELLTKEQFDKLLQAPNTTTLARLLHQSVYHLTVDDLNDLDRLESILMAE 65
           ELN  I  +E EL++K+ F++++Q  +  +L  +L  ++Y   + D  D D  E+ L  E
Sbjct: 5   ELNPLIRGRELELISKDTFEQMIQTDSIDSLGEILQSTIYQPYIYDGFDKD-FEANLSQE 63

Query: 66  LTKTYRWAFAETPQPDIVQLFTLRYTYHNVKVLLKAKASQADLSHLLLPIGDKPLVALEH 125
            +K ++W     P+P+IV ++T+RYT+HN+KVL KA+ +  +L HL +  G   L  L+
Sbjct: 64  RSKLFQWLKESAPEPEIVWIYTMRYTFHNLKVLTKAEITGQNLDHLYIHDGFYSLEVLKD 123

Query: 126 LIRTMTSDEFPKEVVTEIQSIWAEYQDYQDIRVLEIGTDLAYFKALKQIAQRLEDPVFQQ 185
            I T  S E P  ++  I+ +    ++   ++ +++  D  +    +++ ++L  P   +
Sbjct: 124 AIHTQVSVELPDSLMDYIREVHEYCEESTILQGIDVIYDRCFLTEQRRLGEQLGYPELLE 183

Query: 186 AVLIVIDLYNLITVRRAKSQNKPISFMMQLLSDEASRPSKTFITLEDDKDLMTWFENVTP 245
            ++  IDL N+ T  R   Q++   FM  ++S   S P  T ++     ++ ++ + +
Sbjct: 184 EIIAFIDLTNITTTARGILQHRSAGFMTTVISSSGSIPKDTLLSFVRG-EMASFTQFLLT 242

Query: 246 DSYMTALKPYSEKLRQGTLQTTELEYLVDECLYHLFAKAKYQVDGPYVLARFLLAKSFEV 305
             Y   LK   + + +  +    LE L D+ L   +  A+ Q  GP  L  FL AK  E
Sbjct: 243 TDYSELLK---QVIHEEQIDLVSLEQLKDDYLSSFYQVAQTQAFGPLPLLAFLNAKEVES 299

Query: 306 KNLRLLAAALANDLPKERVIERMRPI 331
           KNLRLL     N    E++ ERMR +
Sbjct: 300 KNLRLLIIGKRNHFSLEQLKERMRQV 325
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2406

A DNA sequence (GASx151) was identified in *S.pyogenes* <SEQ ID 7311> which encodes the amino acid sequence <SEQ ID 7312>. Analysis of this protein sequence reveals the following:

```
Possible site: 29

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0484(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAA04274 GB:D17462 Na+ -ATPase subunit G [Enterococcus hirae]
 Identities = 45/101 (44%), Positives = 65/101 (63%)

Query: 6   YKVGVIGNRDVILPFQMIGFQTFPVIKPQDAINQLRQLAMEDFGIIYITEDIAAAIPEAL 65
           YK+GV+G++D + PF++ GF         +    + ++A  ++G+IYITE  A  +PE +
Sbjct: 3   YKIGVVGDKDSVSPFRLFGFDVQHGTTKTEIRKTIDEMAKNEYGVIYITEQCANLVPETI 62

Query: 66  THYDNQVLPAVIPLPTHQGAQGIGLSRIQAMVEKAVGQNIL 106
             Y  Q+ PA+I +P+HQG  GIGL  IQ  VEKAVGQNIL
Sbjct: 63  ERYKGQLTPAIILIPSHQGTLGIGLEEIQNSVEKAVGQNIL 103
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2407**

A DNA sequence (GASx152R) was identified in *S.pyogenes* <SEQ ID 7313> which encodes the amino acid sequence <SEQ ID 7314>. Analysis of this protein sequence reveals the following:

```
Possible site: 21

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1048(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2408**

A DNA sequence (GASx156) was identified in *S.pyogenes* <SEQ ID 7315> which encodes the amino acid sequence <SEQ ID 7316>:

```
EYSIIPQLKETIHYIELKLEEAERASLVRIMKITS
```

Analysis of this protein sequence reveals the following:

```
Possible site: 16

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5026(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAA04277 GB:D17462 Na+ -ATPase subunit D [Enterococcus hirae]
 Identities = 119/201 (59%), Positives = 151/201 (74%), Gaps = 2/201 (0%)

Query: 10  RLNVKPTRMELSNLKNRLKTATRGHKLLKDKRDELMRRFVDLIRENNELRQTIEKELAAN 69
           RLNV PTRMEL+ LK +L TATRGHKLLKDK+DELMR+F+ LIR+NNELRQ IEKE
Sbjct: 2   RLNVNPTRMELTRLKKQLTTATRGHKLLKDKQDELMRQFILLIRKNNELRQAIEKETQTA 61

Query: 70  MKEFVLAKASENSLMVEELFAVPVHEVTLWIDIENIMSVNVPKFHVQSNTAREQEQGEFA 129
           MK+FVLAK++     ++EL A+P   V++ +  +NIMSV VP  + Q +    +   E
Sbjct: 62  MKDFVLAKSTVEEAFIDELLALPAENVSISVVEKNIMSVKVPLMNFQYDETLNETPLE-- 119

Query: 130 YSYLSSNSEMDNTIQKTKELLEKLLRLAEVEKTCQLMADDIEKTRRRVNGLEYSIIPQLK 189
           Y YL SN+E+D +I    +LL KLL+LAEVEKTCQLMA++IEKTRRRVN LEY  IPQL+
Sbjct: 120 YGYLHSNAELDRSIDGFTQLLPKLLKLAEVEKTCQLMAEEIEKTRRRVNALEYMTIPQLE 179

Query: 190 ETIHYIELKLEEAERASLVRI 210
           ETI+YI++KLEE ERA + R+
```

```
Sbjct: 180 ETIYYIKMKLEENERAEVTRL 200
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2409**

A DNA sequence (GASx161R) was identified in *S.pyogenes* <SEQ ID 7317> which encodes the amino acid sequence <SEQ ID 7318>. Analysis of this protein sequence reveals the following:

```
Possible site: 27

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2410**

A DNA sequence (GASx164) was identified in *S.pyogenes* <SEQ ID 7319> which encodes the amino acid sequence <SEQ ID 7320>. Analysis of this protein sequence reveals the following:

```
Possible site: 36

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.06    Transmembrane     9 -  25 (   9 -  25)

----- Final Results -----
               bacterial membrane --- Certainty=0.1426(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related sequence was also identified <SEQ ID 9091> which encodes the amino acid sequence <SEQ ID 9092>. Analysis of this protein sequence reveals the following:'

```
     Possible cleavage site: 33
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty= 0.300(Affirmative) < succ>
               bacterial membrane --- Certainty= 0.000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty= 0.000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2411

A DNA sequence (GASx165) was identified in *S.pyogenes* <SEQ ID 7321> which encodes the amino acid sequence <SEQ ID 7322>. Analysis of this protein sequence reveals the following:

```
Possible site: 59

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2251(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2412

A DNA sequence (GASx166) was identified in *S.pyogenes* <SEQ ID 7323> which encodes the amino acid sequence <SEQ ID 7324>. Analysis of this protein sequence reveals the following:

```
Possible site: 34

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2413

A DNA sequence (GASx167) was identified in *S.pyogenes* <SEQ ID 7325> which encodes the amino acid sequence <SEQ ID 7326>. Analysis of this protein sequence reveals the following:

```
Possible site: 31

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2414**

A DNA sequence (GASx168R) was identified in *S.pyogenes* <SEQ ID 7327> which encodes the amino acid sequence <SEQ ID 7328>. Analysis of this protein sequence reveals the following:

```
Possible site: 22

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2415**

A DNA sequence (GASx169R) was identified in *S.pyogenes* <SEQ ID 7329> which encodes the amino acid sequence <SEQ ID 7330>. Analysis of this protein sequence reveals the following:

```
Possible site: 31

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2416**

A DNA sequence (GASx170) was identified in *S.pyogenes* <SEQ ID 7331> which encodes the amino acid sequence <SEQ ID 7332>. Analysis of this protein sequence reveals the following:

```
Possible site: 61

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.34   Transmembrane  154 - 170 ( 153 - 170)
   INTEGRAL    Likelihood = -1.12   Transmembrane   20 -  36 (  19 -  36)
   INTEGRAL    Likelihood = -0.69   Transmembrane   52 -  68 (  52 -  68)
   INTEGRAL    Likelihood = -0.53   Transmembrane  399 - 415 ( 399 - 415)

----- Final Results -----
               bacterial membrane --- Certainty=0.1935(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB05347 GB:AP001512 cystathionine beta-lyase [Bacillus halodurans]
 Identities = 200/384 (52%), Positives = 262/384 (68%), Gaps = 3/384 (0%)

Query: 79  IAEVYEMRENTTLLHGYTVIDEFTGAASVPIYQTSTFHNSELYCPSQKHLYTRFSNPTTE 138
            ++E Y ++  T LLH    +D+ TGA SVPI    STFH  + +     + Y+R  NPT +
Sbjct: 1   MSEQYSLQ--TKLLHNEHKVDQATGAVSVPIQHASTFHQFD-FDTFGTYDYSRSGNPTRD 57

Query: 139 ALEDGLACLEKATYAVAYASGMAAISTVLMLLKAGDHVIFPLEVYGGTCQFATAILPNYQ 198
            ALE  +A LE   +  A+ASGMAAIST  MLL  GDHV+   +VYGGT +  T +L
Sbjct: 58  ALEAAIAELEGGNHGFAFASGMAAISTAFMLLSKGDHVVLTKDVYGGTFRLVTEVLTRLG 117

Query: 199 IETSFVDMADLATVKASIRPNTRMIYLETPSNPLLKICDISELVQLAKAYGVLTVADNTF 258
            IE +FVDM +LA V A+IRPNTR++Y+ETPSNP L I DI  +V LAK +  LT  DNTF
Sbjct: 118 IEHTFVDMTNLAEVAAAIRPNTRVLYMETPSNPTLNITDIRGVVSLAKEHECLTFLDNTF 177

Query: 259 MTSLYQEPLAMGVDIVVESVTKFINGHSDVVAGLAATNNEAIYNQLKLFQKNFGAIVGVE 318
            +T   Q PL +GVD+V+ S TKFI GHSDVVAGLA T NE +  +L   Q +FGAI+GV+
Sbjct: 178 LTPALQRPLELGVDVVLHSATKFIGGHSDVVAGLAVTKNEELGKKLAFLQNSFGAILGVQ 237

Query: 319 DAWLILRGMKTMGIRMEQAVKNAQQLANYLAKHPKVLKVHYPGLDSHPNHDTHLQQAKNG 378
            D WL+LRG+KT+ +RME   K AQQ+A +L   P+V +V+YPGL  HP H+   +QA+
Sbjct: 238 DVWLVLRGLKTLHVRMEHGEKGAQQIAEWLQGVPEVKRVYYPGLKDHPGHELQKRQAEGF 297

Query: 379 GAVLSFELASKEELMTFTHRIQLPILAVSLGGVESILSHPATMSHACLSPQARLEQGVVD 438
            GAVLSFEL ++E +  F   ++LP+ AVSLG VESILS+PA MSHA +  + R  +G+ D
Sbjct: 298 GAVLSFELENEEAVRRFVEHVKLPVFAVSLGAVESILSYPAKMSHAAMPKEEREARGIRD 357

Query: 439 GLLRLSCGVENIEDLLADFEQALA 462
            GLLRLS G+E  E+L+ADF+ A A
Sbjct: 358 GLLRLSVGLEKPEELMADFKAAFA 381
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2417**

A DNA sequence (GASx178) was identified in *S.pyogenes* <SEQ ID 7333> which encodes the amino acid sequence <SEQ ID 7334>. Analysis of this protein sequence reveals the following:

```
Possible site: 21

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1492(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2418**

A DNA sequence (GASx182) was identified in *S.pyogenes* <SEQ ID 7335> which encodes the amino acid sequence <SEQ ID 7336>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
```

```
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2584(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2419**

A DNA sequence (GASx187) was identified in *S.pyogenes* <SEQ ID 7337> which encodes the amino acid sequence <SEQ ID 7338>. Analysis of this protein sequence reveals the following:

```
Possible site: 61

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2084(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2420**

A DNA sequence (GASx188) was identified in *S.pyogenes* <SEQ ID 7339> which encodes the amino acid sequence <SEQ ID 7340>. Analysis of this protein sequence reveals the following:

```
Possible site: 34

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2060(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAG05515 GB:AE004640 conserved hypothetical protein [Pseudomonas aeruginosa]
 Identities = 140/442 (31%), Positives = 208/442 (46%), Gaps = 73/442 (16%)

Query: 2   KKYLNQNVYDALIERLHFLFNDFPIVYISFSGGKDSGLLLNILLDFRDKYYPDREIG--- 58
           K Y + +V+ A + RL  +F +F  V ++FSGGKDS + L + LD        RE+G
Sbjct: 4   KHYQDADVHAATLSRLRLVFRNFERVCVAFSGGKDSSVTLQLALDVA------RELGRSP 57

Query: 59  --VFHQDFEAQYSLTTKYVQETFTSLEGRKKVSLYWVCLPMATRTALSSYEMFWYPWDDK 116
             V   D E QY  T  +V E      GR  V  +WVCLP+  R A S  E +W  W+
Sbjct: 58  VDVLFIDLEGQYQATIDHVSEML----GRPDVRPWWVCLPLNLRNASSLEEPYWCCWEPG 113

Query: 117 TEDIWVRPMPSQDYVINLENNSITTYRYKMNQEDLAKQFGRWYKQIHGNQKTVCILGNRA 176
```

```
            E  WVRP+P Q  VI+ +      YRY+M  E+    F  W  +    + T  ++G R+
Sbjct: 114 AEADWVRPLPKQRGVIS-DPAFFPFYRYRMEFEEFVAGFNAWLAR---EEPTAFLVGIRS 169

Query: 177 SESLHRYSGFINKKYGYQKEC------------WITKQFKDVWTAS--PLYDWSVEDIWH 222
            ESL+RY     K+    K+C            W  +   +    S  P+YDW  ED+W
Sbjct: 170 DESLNRYLAV--KRRSRAKQCAWTPPGGSAPLAWSARDRANPQAVSFFPIYDWRFEDLWR 227

Query: 223 AYYKFSYSYNELYDLFYKAGLKPSQMRVASPFQDYAVDSLNLYRIIDQETWVKLLGRVQG 282
                  Y+YN LYD  Y+AG+  SQMR+  P+ D     L+L+  I+  TW K++ RV G
Sbjct: 228 CVADHGYAYNRLYDQMYRAGVPFSQMRICQPYGDDQRKGLDLFHRIEPRTWFKVVRRVAG 287

Query: 283 VNFSNIYGRTKAMGYK-SIALPKGH-SWKSYTQFLLSTLPVRLRNNYVRKFNKSIDFWHK 340
             N+   Y R + +GY+  + LP    +W+ Y+QFLL ++P  LR  Y R+  + I +W +
Sbjct: 288 ANYGARYCRQRFLGYRGGLGLPPSFGTWREYSQFLLRSMPPPLRGIYQRRIERFILWWKQ 347

Query: 341 TGGGLAEETINELIEKGYRIARNGISNYTSFKHSRVIFLDQ-IPDDTDDIVTTKDIPSWK 399
                LA                              I+ D  IP     +   +  PSW+
Sbjct: 348 HDYPLA------------------------------IWPDAGIP----ALENRRKQPSWR 373

Query: 400 RMCFCILKNDHICRTMGFGLTR 421
            R+   +LK D + R++ FG ++
Sbjct: 374 RIALSLLKQD-MARSLSFGFSQ 394
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2421**

A DNA sequence (GASx189) was identified in *S.pyogenes* <SEQ ID 7341> which encodes the amino acid sequence <SEQ ID 7342>. Analysis of this protein sequence reveals the following:

```
Possible site: 45

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4121(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC73702 GB:AE000165 orf, hypothetical protein [Escherichia
           coli]
 Identities = 79/162 (48%), Positives = 110/162 (67%), Gaps = 1/162 (0%)

Query: 7   PVYEIKSIPIEKISPNDYNPNSVAPPEMKLLYDSIKSDGYTMPIVCYYDKEEDRYSIVDG 66
           PV  +  +   ++ PNDYNPN+VAPPE KLL  SI+ DG+T PIV  +  +++   IVDG
Sbjct: 46  PVDCVLWVKNSQLMPNDYNPNNVAPPEKKLLQKSIEIDGFTQPIVVTHT-DKNAMEIVDG 104

Query: 67  FHRYRIMLDYSDIYERESGRLPVSVIDKSLDYRMASTIRHNRARGSHDVDLMSQIVKDLH 126
           FHR+ I    S +  R  G LPV+ ++ + + R+A+TIRHNRARG H +  MS+IV++L
Sbjct: 105 FHRHEIGKGSSSLKLRLKGYLPVTCLEGTRNQRIAATIRHNRARGRHQITAMSEIVRELS 164

Query: 127 ECGRSDNWIAKHLGMDKDEILRLKQITGLASLFKDHEFNQSW 168
           + G  DN I K LGMD DE+LRLKQI GL  LF D +++++W
Sbjct: 165 QLGWDDNKIGKELGMDSDEVLRLKQINGLQELFADRQYSRAW 206
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2422**

A repeated DNA sequence (GASx192R) was identified in *S.pyogenes* <SEQ ID 7343> which encodes the amino acid sequence <SEQ ID 7344>. Analysis of this protein sequence reveals the following:

```
Possible site: 13

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4301(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA63509 GB:X92946 transposase [Lactococcus lactis]
 Identities = 23/36 (63%), Positives = 28/36 (76%)

Query: 1   MQDKLVTEAFNQAYNREKPKEGVIVHTDQGSQYTGA 36
           MQDKLV + F QA  +E P+ G+IVHTDQGSQYT +
Sbjct: 134 MQDKLVRDCFLQACGKEHPQPGLIVHTDQGSQYTSS 169
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2423**

A DNA sequence (GASx194R) was identified in *S.pyogenes* <SEQ ID 7345> which encodes the amino acid sequence <SEQ ID 7346>. Analysis of this protein sequence reveals the following:

```
Possible site: 26

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA63508 GB:X92946 hypothetical protein [Lactococcus lactis]
 Identities = 64/96 (66%), Positives = 78/96 (80%)

Query: 1  MPRKTFDKAFKLSAVKLILEEEQSVKMVSSTLEIHPNSLYQWIQEYEKYGESAFPGHGSA 60
          M R+ FDK FK SAVKLILEE  SVK VS   LE+H NSLY+W+QE E+YGESAFPG+G+A
Sbjct: 1  MARRKFDKQFKNSAVKLILEEGYSVKEVSQELEVHANSLYRWVQEVEEYGESAFPGNGTA 60

Query: 61 LRHAQFETKKLEKEHKLLQEELALLKKFQVFLKPNR 96
          L +AQ + K LEKE++ LQEEL LLKKF+VFLK ++
Sbjct: 61 LANAQHKIKLLEKENRYLQEELELLKKFRVFLKRSK 96
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2424**

A DNA sequence (GASx195R) was identified in *S.pyogenes* <SEQ ID 7347> which encodes the amino acid sequence <SEQ ID 7348>. Analysis of this protein sequence reveals the following:

```
Possible site: 13

>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-11.30   Transmembrane  179 - 195 ( 173 - 201)
   INTEGRAL    Likelihood = -8.86   Transmembrane  229 - 245 ( 224 - 254)
   INTEGRAL    Likelihood = -8.39   Transmembrane  289 - 305 ( 280 - 307)
   INTEGRAL    Likelihood = -8.23   Transmembrane  417 - 433 ( 410 - 435)
   INTEGRAL    Likelihood = -5.89   Transmembrane  324 - 340 ( 323 - 349)
   INTEGRAL    Likelihood = -4.73   Transmembrane  260 - 276 ( 256 - 278)
   INTEGRAL    Likelihood = -4.51   Transmembrane   96 - 112 (  91 - 113)
   INTEGRAL    Likelihood = -4.25   Transmembrane   24 -  40 (  20 -  43)
   INTEGRAL    Likelihood = -2.44   Transmembrane  344 - 360 ( 342 - 360)

----- Final Results -----
               bacterial membrane --- Certainty=0.5522(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB75191 GB:AL139075 putative integral membrane protein
           [Campylobacter jejuni]
 Identities = 177/430 (41%), Positives = 274/430 (63%), Gaps = 8/430 (1%)

Query: 5   IIISAIALAIGIGYRTKINIGLLAIAFSYLIATTLMGLSPKELLHFWPTSLFFTIFSVSL 64
           +IIS+I +AI +GY T+ N+G+ A+ F+Y+I    M L+PK+++ FWP S+FF IF+VSL
Sbjct: 6   LIISSIIVAIILGYITRHNVGIFAMIFAYIIGAFFMDLAPKKIIAFWPISIFFVIFAVSL 65

Query: 65  FYNVATTNGTLDVLAQHILYRTRTHPNALYMILYLIATLLSALGAGFFTTMAVCCPLAIT 124
           FYN AT NGTL+ LA H++YR   HP  L  ++++++ +++ALGAGF+T +A   PL
Sbjct: 66  FYNFATVNGTLEKLAGHLMYRFANHPYLLPFVIFVVSAIIAALGAGFYTVLAFMAPLTFL 125

Query: 125 LCQKADKHPLIGAQAVNWGASGGANLITSGSGIVFQGLFKQMGWE-EQAFSLGNHIFIVS 183
           LC K     + GA A+N+GA GGAN ITS SGI+F+GL +  G E  +AF+  + IF  +
Sbjct: 126 LCDKIGLSKIAGAMAINYGALGGANFITSQSGIIFRGLMENSGIEANEAFANSSIIFAFT 185

Query: 184 IIYPLIVLLLLSCYIRYSKGRTNSSLT-IDQPPVLSKVQRQTTLLMISSMVLVWLFPLLL 242
           II P++VL     +  ++  + N  ++ I +P      Q+ T +LM   +V+V +FP+L
Sbjct: 186 IILPIVVL----SFFVFNAFKNNIKISVISKPDPFDYKQKTTLILMFMMIVVVLIFPVLN 241

Query: 243 LIFPNIAWIATYRQTFDIGFVSILMVCLALRLKLGKQEAILAKVPWAIIIMLCGMSLLMS 302
           +IFP+   I+ + +  DI  ++++ V +AL LKL  ++ ++A +PW  +IM+CG+ +L+S
Sbjct: 242 IIFPHNETISYFNKKIDIAMIAMIFVAIALFLKLADEKQVVALIPWGTLIMICGVGMLIS 301

Query: 303 LAVKSGLVTLIGHLITTTIPHFWLPLFFCVIAGVMSLFSSTLSVVAPTLFPIIATISAQS 362
           +AV++G + L   L+   I   ++PL  C IA  MSLFSSTL VV P LFPI+ +I+A S
Sbjct: 302 IAVEAGAIKLFSDLVENEINVIFIPLIMCAIAAFMSLFSSTLGVVTPALFPIVPSIAASS 361

Query: 363 PHIDIRLLTTATIIGALSTNISPFSSAGSLIQLSLPHIEERSLAFKKQILLGVPISLSLA 422
             +   LL +  ++GA ++ ISPFSS GSLI  S P   +  L FK  ++  VPI    A
Sbjct: 362 -GLSEALLFSCIVVGAQASAISPFSSGGSLILGSCPDKYKEKL-FKDLLIKAVPIGFIAA 419

Query: 423 LLTIWILMLL 432
           +L   I+  +
Sbjct: 420 ILATIIMSFI 429
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

## Example 2425

A DNA sequence (GASx196) was identified in *S.pyogenes* <SEQ ID 7349> which encodes the amino acid sequence <SEQ ID 7350>. Analysis of this protein sequence reveals the following:

```
Possible site: 57

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0563(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC45128 GB:U65510 nicotinate-nucleotide pyrophosphorylase
          [Rhodospirillum rubrum]
 Identities = 116/277 (41%), Positives = 170/277 (60%), Gaps = 4/277 (1%)

Query: 17  LTPFQIDDTLKAALREDV-HSEDYSTNAIFDHHGQAKVSLFAKEAGVLAGLTVFQRVFTL 75
           L+PF ID+ ++ AL ED+  + D ++ A      +A     A++ G+LAGL   +  F L
Sbjct: 10  LSPFAIDEAVRRALAEDLGRAGDITSTATIPAATRAHARFVARQPGILAGLGCARSAFAL 69

Query: 76  FDTEVTFQNPHQFKDGDRLTSGDLVLEIIGSVRSLLTCERVALNFLQHLSGIASMTAAYV 135
            D  VTF  P   +DG  + +G  V E+ G+ R++L  ER ALNFL HLSGIA+ T  +
Sbjct: 70  LDDTVTFTTP--LEDGAEIAAGQTVAEVAGAARTILAAERTALNFLGHLSGIATRTRRFG 127

Query: 136 EALGDDRIKVFDTRKTTPNLRLFEKYAVRVGGGYNHRFNLSDAIMLKDNHIAAVGSVQKA 195
           +A+   R ++  TRKTTP LR  EKYAVR GGG NHRF L DA+++KDNHIA  G V  A
Sbjct: 128 DAIAHTRARLTCTRKTTPGLRGLEKYAVRCGGGSNHRFGLDDAVLIKDNHIAVAGGVSAA 187

Query: 196 IAQARAYAPFVKMVEVEVESL-AAAEEAAAAGVDIIMLDNMSLEQIEQAITLIAGRSRIE 254
           +++ARA    +  +E+EV++L   AE  A  G ++++LDNM    + +A+ ++AGR   E
Sbjct: 188 LSRARAGVGHMVRIEIEVDTLEQLAEVLAVGGAEVVLLDNMDAPTLTRAVDMVAGRLVTE 247

Query: 255 CSGNIDMTTISRFRGLAIDYVSSGSLTHSAKSLDFSM 291
            SG + + TI+      +DY+S G+LTHS  +LD  +
Sbjct: 248 ASGGVSLDTIAALAESGVDYISVGALTHSVTTLDIGL 284
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

## Example 2426

A DNA sequence (GASx199) was identified in *S.pyogenes* <SEQ ID 7351> which encodes the amino acid sequence <SEQ ID 7352>. Analysis of this protein sequence reveals the following:

```
Possible site: 25

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1649(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2427**

A DNA sequence (GASx201) was identified in *S.pyogenes* <SEQ ID 7353> which encodes the amino acid sequence <SEQ ID 7354>. Analysis of this protein sequence reveals the following:

```
Possible site: 19

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2428**

A DNA sequence (GASx203) was identified in *S.pyogenes* <SEQ ID 7355> which encodes the amino acid sequence <SEQ ID 7356>. Analysis of this protein sequence reveals the following:

```
Possible site: 37

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2429**

A DNA sequence (GASx210) was identified in *S.pyogenes* <SEQ ID 7357> which encodes the amino acid sequence <SEQ ID 7358>. Analysis of this protein sequence reveals the following:

```
Possible site: 29

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2430**

A DNA sequence (GASx211) was identified in *S.pyogenes* <SEQ ID 7359> which encodes the amino acid sequence <SEQ ID 7360>. Analysis of this protein sequence reveals the following:

```
Possible site: 24

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
              bacterial outside --- Certainty=0.3000(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2431**

A DNA sequence (GASx213) was identified in *S.pyogenes* <SEQ ID 7361> which encodes the amino acid sequence <SEQ ID 7362>. Analysis of this protein sequence reveals the following:

```
Possible site: 14

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.4430(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2432**

A DNA sequence (GASx219) was identified in *S.pyogenes* <SEQ ID 7363> which encodes the amino acid sequence <SEQ ID 7364>. Analysis of this protein sequence reveals the following:

```
Possible site: 15

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2433

A DNA sequence (GASx220) was identified in *S.pyogenes* <SEQ ID 7365> which encodes the amino acid sequence <SEQ ID 7366>. Analysis of this protein sequence reveals the following:

```
Possible site: 24

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0530(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2434

A DNA sequence (GASx231R) was identified in *S.pyogenes* <SEQ ID 7367> which encodes the amino acid sequence <SEQ ID 7368>. Analysis of this protein sequence reveals the following:

```
Possible site: 30

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2435

A DNA sequence (GASx237) was identified in *S.pyogenes* <SEQ ID 7369> which encodes the amino acid sequence <SEQ ID 7370>. Analysis of this protein sequence reveals the following:

```
Possible site: 52

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4961(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB49143 GB:AJ248283 hypothetical protein [Pyrococcus abyssi]
 Identities = 79/229 (34%), Positives = 131/229 (56%), Gaps = 11/229 (4%)

Query: 18  MRFTIDQNMQFPLVEIDLEHGGSVYLQQGSMVYHTENVTLNTKLNGKGSGLGKLVGAIGR 77
            M + I+    F L+E++L  G +V  + G+MVY    V++ TK  G       L+GA+ R
Sbjct: 1   MEYRIEHRPSFSLLEVNLREGEAVQAEAGAMVYMDPTVSIETKARGG------LLGALKR 54

Query: 78  SMVSGESMFITQAMSNGDGKLALAPNTPGQIVALELGEKQYRLNDGAFLALDGSAQYKME 137
            S++ GES F+   +  G G++  AP  PG I++LEL    Y    GAFL         ++
Sbjct: 55  SVLGGESFFMN--VFRGPGRVGFAPGYPGDIISLELNGTLYA-QSGAFLVASEGIDIDVK 111

Query: 138 RQNIGKALFGGQGGLFVMTTEGLGTLLANSFGSIKKITLDGGTMTIDNAHVVAWSRELDY 197
                GK +FG +G +F++  +G G +  +S+G+I+KITL G ++ +D  H+VA++  +D+
Sbjct: 112 FGG-GKTIFGREG-VFLLELKGKGIVFLSSYGAIEKITLRGESVIVDTGHMVAFTEGIDF 169

Query: 198 DIHLENGFMQSIGTGEGVVNTFRGHGEIYIQSLNLEQFAGTLKRYLPTS 246
             I    G   ++ +GEG+V  F GHG++YIQ+ +L+ F   +  +LP S
Sbjct: 170 RIRKIGGLKATLFSGEGLVFEFSGHGDVYIQTRSLDGFLSWILPHLPKS 218
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2436**

A DNA sequence (GASx240R) was identified in *S.pyogenes* <SEQ ID 7371> which encodes the amino acid sequence <SEQ ID 7372>. Analysis of this protein sequence reveals the following:

```
Possible site: 35

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2745(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2437**

A DNA sequence (GASx241) was identified in *S.pyogenes* <SEQ ID 7373> which encodes the amino acid sequence <SEQ ID 7374>. Analysis of this protein sequence reveals the following:

```
Possible site: 21

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-10.14   Transmembrane  196 - 212 ( 187 - 215)
   INTEGRAL    Likelihood = -8.01   Transmembrane  160 - 176 ( 156 - 179)
   INTEGRAL    Likelihood = -5.89   Transmembrane  116 - 132 ( 110 - 134)
   INTEGRAL    Likelihood = -4.57   Transmembrane   74 -  90 (  73 -  97)
   INTEGRAL    Likelihood = -2.66   Transmembrane   51 -  67 (  50 -  68)
   INTEGRAL    Likelihood = -2.60   Transmembrane    8 -  24 (   7 -  27)
   INTEGRAL    Likelihood = -1.28   Transmembrane  344 - 360 ( 344 - 360)
   INTEGRAL    Likelihood = -0.22   Transmembrane   30 -  46 (  30 -  46)
```

```
----- Final Results -----
               bacterial membrane --- Certainty=0.5055(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAC10175 GB:AJ278302 histidine kinase [Streptococcus pneumoniae]
 Identities = 136/449 (30%), Positives = 234/449 (51%), Gaps = 26/449 (5%)

Query: 8   FLLLSIIVYYMTKIYIFSFLSDITLP---VWKQLTI-LALALFFNQFPYLS-----PLLI 58
           ++LL  +V +  KI IF  +  I+L    ++K     + LA+ F    Y+         +
Sbjct: 5   WILLYTLVTHGLKIVIFFKVDGISLTFERIFKAFLFKILLAVVFGMLGYMVGNVYLSYFM 64

Query: 59  DPL----LFLVVLRQETKQLFSLKALFLAVAPSVLVDLLSRFMGTIVIPYLFLSSGIYLG 114
           +PL    L  ++LR+  K+L     LF  + P +LV+L  R +   V+P  FL  G
Sbjct: 65  EPLYGIGLSFLLLRELPKKLL----LFYGLFPMILVNLFYRGVSYFVLP--FLGQGQVYD 118

Query: 115 HIIFDLLAYLLIFPSFAIINYMIGKDYKMIC-QSGYSKRSHNFYQTLLMFVLVYYVDIFV 173
              F  L  ++IF  F  + ++   DY     + G   +      T + +++  Y  +
Sbjct: 119 DYSFIWLC-IIIFNFFISLAFLKWLDYDFTSLRKGILDKDFQKSLTQINWIMGAYYLVIQ 177

Query: 174 ILGFTDPFLHFHHSLFVPTPYKLLFLMFILLLVYLLSYFNHSSKEYLKNELRREQQAYMT 233
            L + +    +   +   T   L+ + ++L  + ++   +   K+ L   L +EQ
Sbjct: 178 NLSYFE----YEQGIQSTTVRHLILVFYLLFFMGIIKKLDTYLKDKLHERLNQEQDLRYR 233

Query: 234 NLEIYGKHLEKLYRDVRAFQSDYLSRIERLGQAIKSESITQIQDIYAQTVHEANDYWDDK 293
            +E Y +H+E+LY++VR+F+ DY + +  L   I+ E + QI++IY   + ++++   D
Sbjct: 234 EMERYSRHIEELYKEVRSFRHDYTNLLTSLRLGIEEEDMEQIKEIYDSVLKDSSEKLQDN 293

Query: 294 HYNISKLRKINISSIKSLLSAKIISAEKSGIDLNVEVPDNIKETYIPELDLLLLMSIFCD 353
            Y++ +L  +   ++KSLL+ K I A    I  NVEVP+ I+   +  LD L ++SI CD
Sbjct: 294 KYDLGRLVNVRDRALKSLLAGKFIKARDKNIVFNVEVPEEIQVEGVSLLDFLTVVSILCD 353

Query: 354 NAIEAALEAQQPHMSIAYFLLGDYQMFVVTNTTKKK-VDINKIFEEGYSSKGSERGIGLS 412
           NAIEA++EA QPH+SIA+F  G  + F++ N+ K++ +DI++IF  G SSKG ERG+GL
Sbjct: 354 NAIEASVEACQPHVSIAFFKNGAQETFIIENSIKEEGIDISEIFSFGASSKGEERGVGLY 413

Query: 413 NAQRILKKYPYLSLRTKSFDKEFSQTLTM 441
              +I++ +P  SL T   D  F Q LT+
Sbjct: 414 TVMKIVESHPNTSLNTTCQDHVFRQVLTV 442
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

## Example 2438

A DNA sequence (GASx242R) was identified in *S.pyogenes* <SEQ ID 7375> which encodes the amino acid sequence <SEQ ID 7376>. Analysis of this protein sequence reveals the following:

```
Possible site: 26

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4165(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2439**

A DNA sequence (GASx243) was identified in *S.pyogenes* <SEQ ID 7377> which encodes the amino acid sequence <SEQ ID 7378>. Analysis of this protein sequence reveals the following:

```
Possible site: 26

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-11.09   Transmembrane  188 - 204 ( 182 - 208)
   INTEGRAL    Likelihood = -7.17   Transmembrane   52 -  68 (  47 -  69)
   INTEGRAL    Likelihood = -4.73   Transmembrane  119 - 135 ( 114 - 142)
   INTEGRAL    Likelihood = -4.62   Transmembrane   83 -  99 (  77 - 107)
   INTEGRAL    Likelihood = -1.86   Transmembrane  328 - 344 ( 328 - 345)
   INTEGRAL    Likelihood = -1.65   Transmembrane    7 -  23 (   6 -  23)
   INTEGRAL    Likelihood = -0.22   Transmembrane   35 -  51 (  35 -  51)

----- Final Results -----
               bacterial membrane --- Certainty=0.5437(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAC10175 GB:AJ278302 histidine kinase [Streptococcus pneumoniae]
 Identities = 123/438 (28%), Positives = 229/438 (52%), Gaps = 49/438 (11%)

Query: 20  VIFAKVSAIKLSWKRVS-------IIGISFVIANMIFDKVIL---IDQLFFIIVSLL--- 66
            VIF KV  I L+++R+        ++ + F +   +   V L   ++ L+ I +S L
Sbjct: 19  VIFFKVDGISLTFERIFKAFLFKILLAVVFGMLGYMVGNVYLSYFMEPLYGIGLSFLLLR 78

Query: 67  SAPKKKLFEHMFNGFFTILIVELLFRVIGSFFLPAVLGFSIGQINNNLKLLELCYLFVLP 126
              PKK L   +F G F +++V L +R +  F LP +     GQ+ ++   + LC + +
Sbjct: 79  ELPKKLL---LFYGLFPMILVNLFYRGVSYFVLPFL---GQGQVYDDYSFIWLC-IIIFN 131

Query: 127 IFYLFSYIFSIDL---SLIRFISEDKMKKWVFWMNTAMFSYYFFAHFLVTVQSGFLALYF 183
             F   +++  +D    SL + I +   +K +  +N  M +YY     L          YF
Sbjct: 132 FFISLAFLKWLDYDFTSLRKGILDKDFQKSLTQINWIMGAYYLVIQNLS---------YF 182

Query: 184 QY---------RSILVFIYLAIFIWVIVKLDRFAKDQLSQKLTQAQNERIAYLENYNQSI 234
            +Y         R +++  YL  F+ +I KLD + KD+L ++L Q Q+ R   +E Y++ I
Sbjct: 183 EYEQGIQSTTVRHLILVFYLLFFMGIIKKLDTYLKDKLHERLNQEQDLRYREMERYSRHI 242

Query: 235 EQLYREIRTVKHDSENILISLKDSIDSGDIDLITRVYDTVIQQSATSMMRTNYEISSLDN 294
            E+LY+E+R+ +HD  N+L SL+  I+  D++ I  +YD+V++ S+  +    Y++  L N
Sbjct: 243 EELYKEVRSFRHDYTNLLTSLRLGIEEEDMEQIKEIYDSVLKDSSEKLQDNKYDLGRLVN 302

Query: 295 IKEAVIRSIMNSKLLEAQYLGIELYIEIPDVIDHLPIKLIDLIVLFTGLVDNAIETAKGS 354
            +++  ++S++  K ++A+   I   +E+P+ I    + L+D + + + L DNAIE +  +
Sbjct: 303 VRDRALKSLLAGKFIKARDKNIVFNVEVPEEIQVEGVSLLDFLTVVSILCDNAIEASVEA 362

Query: 355 RRPFLSIAYFKQDNKQLFIIENSTKTNRVDIAKRFDAQQQNSAH--------FLTVLDSY 406
             +P +SIA+FK   ++ FIIENS K   +DI++ F     +            + +++S+
Sbjct: 363 CQPHVSIAFFKNGAQETFIIENSIKEEGIDISEIFSFGASSKGEERGVGLYTVMKIVESH 422

Query: 407 PQITLSTKSDHYRLRQLL 424
            P  +L+T    +  RQ+L
Sbjct: 423 PNTSLNTTCQDHVFRQVL 440
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2440**

A DNA sequence (GASx248) was identified in *S.pyogenes* <SEQ ID 7379> which encodes the amino acid sequence <SEQ ID 7380>. Analysis of this protein sequence reveals the following:

```
Possible site: 32

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5665(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2441**

A DNA sequence (GASx255) was identified in *S.pyogenes* <SEQ ID 7381> which encodes the amino acid sequence <SEQ ID 7382>. Analysis of this protein sequence reveals the following:

```
Possible site: 19

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1437(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2442**

A DNA sequence (GASx270R) was identified in *S.pyogenes* <SEQ ID 7383> which encodes the amino acid sequence <SEQ ID 7384>. Analysis of this protein sequence reveals the following:

```
Possible site: 21

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -5.89   Transmembrane   20 -  36 (  17 -  36)

----- Final Results -----
               bacterial membrane --- Certainty=0.3357(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2443**

A DNA sequence (GASx272) was identified in *S.pyogenes* <SEQ ID 7385> which encodes the amino acid sequence <SEQ ID 7386>. Analysis of this protein sequence reveals the following:

```
Possible site: 58

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2488(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB11887 GB:Z99104 ribosomal protein S7 (BS7) [Bacillus subtilis]
 Identities = 117/156 (75%), Positives = 139/156 (89%)

Query: 1   MSRKNQAPKREVLPDPLYNSKIVTRLINRVMLDGKRGTAATIVYDAFNAIKEATGNDALE 60
           M RK    KR+VLPDP+YNSK+V+RLIN++M+DGK+G   TI+Y +F+ IKE TGNDA+E
Sbjct: 1   MPRKGPVAKRDVLPDPIYNSKLVSRLINKMMIDGKKGKPQTILYKSFDIIKERTGNDAME 60

Query: 61  VFETAMDNIMPVLEVRARRVGGSNYQVPVEVRPERRTTLGLRWLVNASRARGEHTMKDRL 120
           VFE A+ NIMPVLEV+ARRVGG+NYQVPVEVRPERRTTLGLRWLVN +R RGE TM++RL
Sbjct: 61  VFEQALKNIMPVLEVKARRVGGANYQVPVEVRPERRTTLGLRWLVNYARLRGEKTMEERL 120

Query: 121 AKEIMDAANNTGASVKKREDTHKMAEANRAFAHFRW 156
           A EI+DAANNTGA+VKKREDTHKMAEAN+AFAH+RW
Sbjct: 121 ANEILDAANNTGAAVKKREDTHKMAEANKAFAHYRW 156
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2444**

A DNA sequence (GASx274) was identified in *S.pyogenes* <SEQ ID 7387> which encodes the amino acid sequence <SEQ ID 7388>. Analysis of this protein sequence reveals the following:

```
Possible site: 61

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related sequence was also identified in GAS <SEQ ID 9095> which encodes the amino acid sequence <SEQ ID 9096>. Analysis of this protein sequence reveals the following:

```
     Possible cleavage site: 52
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty= 0.291(Affirmative) < succ>
               bacterial membrane --- Certainty= 0.000(Not Clear) < succ>
                bacterial outside --- Certainty= 0.000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2445**

A DNA sequence (GASx275R) was identified in *S.pyogenes* <SEQ ID 7389> which encodes the amino acid sequence <SEQ ID 7390>. Analysis of this protein sequence reveals the following:

```
Possible site: 16

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5664(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2446**

A DNA sequence (GASx283) was identified in *S.pyogenes* <SEQ ID 7391> which encodes the amino acid sequence <SEQ ID 7392>. Analysis of this protein sequence reveals the following:

```
Possible site: 18

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0724(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2447**

A DNA sequence (GASx298) was identified in *S.pyogenes* <SEQ ID 7393> which encodes the amino acid sequence <SEQ ID 7394>. Analysis of this protein sequence reveals the following:

```
Possible site: 25

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2840(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2448**

A DNA sequence (GASx300) was identified in *S.pyogenes* <SEQ ID 7395> which encodes the amino acid sequence <SEQ ID 7396>. Analysis of this protein sequence reveals the following:

```
Possible site: 18

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -1.91   Transmembrane    4 -  20 (   4 -  20)

----- Final Results -----
              bacterial membrane --- Certainty=0.1765(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2449**

A DNA sequence (GASx301) was identified in *S.pyogenes* <SEQ ID 7397> which encodes the amino acid sequence <SEQ ID 7398>. Analysis of this protein sequence reveals the following:

```
Possible site: 33

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.4884(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2450**

A repeated DNA sequence (GASx302) was identified in *S.pyogenes* <SEQ ID 7399> which encodes the amino acid sequence <SEQ ID 7400>. Analysis of this protein sequence reveals the following:

```
Possible site: 22

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
```

```
                  bacterial cytoplasm --- Certainty=0.2581(Affirmative) < succ>
                   bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2451**

A DNA sequence (GASx316) was identified in *S.pyogenes* <SEQ ID 7401> which encodes the amino acid sequence <SEQ ID 7402>. Analysis of this protein sequence reveals the following:

```
Possible site: 18

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.80    Transmembrane   23 -  39 (  22 -  39)

----- Final Results -----
              bacterial membrane --- Certainty=0.1319(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2452**

A DNA sequence (GASx323R) was identified in *S.pyogenes* <SEQ ID 7403> which encodes the amino acid sequence <SEQ ID 7404>. Analysis of this protein sequence reveals the following:

```
Possible site: 28

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.0005(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2453**

A DNA sequence (GASx334) was identified in *S.pyogenes* <SEQ ID 7405> which encodes the amino acid sequence <SEQ ID 7406>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
```

```
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2454

A DNA sequence (GASx336) was identified in *S.pyogenes* <SEQ ID 7407> which encodes the amino acid sequence <SEQ ID 7408>. Analysis of this protein sequence reveals the following:

```
Possible site: 31

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3379(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2455

A DNA sequence (GASx361R) was identified in *S.pyogenes* <SEQ ID 7409> which encodes the amino acid sequence <SEQ ID 7410>. Analysis of this protein sequence reveals the following:

```
Possible site: 22

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2807(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2456

A DNA sequence (GASx387) was identified in *S.pyogenes* <SEQ ID 7411> which encodes the amino acid sequence <SEQ ID 7412>. Analysis of this protein sequence reveals the following:

```
Possible site: 16

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2740(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2457**

A DNA sequence (GASx389) was identified in *S.pyogenes* <SEQ ID 7413> which encodes the amino acid sequence <SEQ ID 7414>. Analysis of this protein sequence reveals the following:

```
Possible site: 21

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0744(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2458**

A DNA sequence (GASx392) was identified in *S.pyogenes* <SEQ ID 7415> which encodes the amino acid sequence <SEQ ID 7416>. Analysis of this protein sequence reveals the following:

```
Possible site: 29

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2162(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2459**

A DNA sequence (GASx393R) was identified in *S.pyogenes* <SEQ ID 7417> which encodes the amino acid sequence <SEQ ID 7418>. Analysis of this protein sequence reveals the following:

```
Possible site: 18

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.2520(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2460**

A DNA sequence (GASx395) was identified in *S.pyogenes* <SEQ ID 7419> which encodes the amino acid sequence <SEQ ID 7420>. Analysis of this protein sequence reveals the following:

```
Possible site: 16

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.2590(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2461**

A DNA sequence (GASx396) was identified in *S.pyogenes* <SEQ ID 7421> which encodes the amino acid sequence <SEQ ID 7422>. Analysis of this protein sequence reveals the following:

```
Possible site: 41

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB13373 GB:Z99111 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 23/88 (26%), Positives = 52/88 (58%)
```

```
Query: 4  KQERIGLVVYLYYNRDARKLSKFGDLYYHSKRSRYLIIYINKNDLDTKLEEMRRLKCVKD 63
          +  R G+VVYL+  + ++ L KFG+++Y SKR +Y+++Y + + ++  ++++     VK
Sbjct: 2  ENRRQGMVVYLHSLKQSKMLRKFGNVHYVSKRLKYVVLYCDMDQIEKTMDKIASYSFVKK 61

Query: 64 IRPSAFDDIDRQFVGNLHRDETNNHQKG 91
          + PS    +  +F   L + +  +++ G
Sbjct: 62 VEPSYKPFLKLEFESKLDKAKEYDYKIG 89
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2462**

A DNA sequence (GASx400) was identified in *S.pyogenes* <SEQ ID 7423> which encodes the amino acid sequence <SEQ ID 7424>. Analysis of this protein sequence reveals the following:

```
Possible site: 13

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2010(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2463**

A DNA sequence (GASx401) was identified in *S.pyogenes* <SEQ ID 7425> which encodes the amino acid sequence <SEQ ID 7426>. Analysis of this protein sequence reveals the following:

```
Possible site: 17

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1176(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2464**

A DNA sequence (GASx402) was identified in *S.pyogenes* <SEQ ID 7427> which encodes the amino acid sequence <SEQ ID 7428>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
```

```
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2938(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2465**

A DNA sequence (GASx403R) was identified in *S.pyogenes* <SEQ ID 7429> which encodes the amino acid sequence <SEQ ID 7430>. Analysis of this protein sequence reveals the following:

```
Possible site: 21

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2466**

A DNA sequence (GASx406) was identified in *S.pyogenes* <SEQ ID 7431> which encodes the amino acid sequence <SEQ ID 7432>. Analysis of this protein sequence reveals the following:

```
Possible site: 31

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-12.26    Transmembrane    15 -  31 (   4 -  36)
   INTEGRAL    Likelihood = -6.64    Transmembrane    96 - 112 (  94 - 115)

----- Final Results -----
               bacterial membrane --- Certainty=0.5904(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2467**

A DNA sequence (GASx408R) was identified in *S.pyogenes* <SEQ ID 7433> which encodes the amino acid sequence <SEQ ID 7434>. Analysis of this protein sequence reveals the following:

```
Possible site: 19

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.23   Transmembrane   17 -  33 (  15 -  34)
   INTEGRAL    Likelihood = -0.85   Transmembrane   38 -  54 (  38 -  54)

----- Final Results -----
               bacterial membrane --- Certainty=0.1893(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2468**

A DNA sequence (GASx412) was identified in *S.pyogenes* <SEQ ID 7435> which encodes the amino acid sequence <SEQ ID 7436>. Analysis of this protein sequence reveals the following:

```
Possible site: 13

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -6.53   Transmembrane    5 -  21 (   4 -  23)

----- Final Results -----
               bacterial membrane --- Certainty=0.3612(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2469**

A DNA sequence (GASx413) was identified in *S.pyogenes* <SEQ ID 7437> which encodes the amino acid sequence <SEQ ID 7438>. Analysis of this protein sequence reveals the following:

```
Possible site: 56

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3422(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA68903 GB:Y07622 lactate oxidase [Streptococcus iniae]
 Identities = 328/392 (83%), Positives = 359/392 (90%), Gaps = 4/392 (1%)

Query: 3   MAQKTVITEETTDFVMDFKTSSAEGNVDFINVFDLEKMAQQVIPKGAFGYIASGAGDTFT 62
            M  K+ +   TT   ++FKTSSAEG+VDF+NVFDLEKMAQ+VIPKGAFGYIASGAGDTFT
Sbjct: 1   MENKSEMINATT---IEFKTSSAEGSVDFVNVFDLEKMAQKVIPKGAFGYIASGAGDTFT 57

Query: 63  LHENIRSFNHKLIVPHSLKGVENPSTEITFDGDYLTSPLILAPVAAHKLANEQGEVASAK 122
            LHENIRSFNHKLI PH LKGVENPSTEITF GD L SP+ILAPVAAHKLANEQGE+ASAK
Sbjct: 58  LHENIRSFNHKLI-PHGLKGVENPSTEITFIGDKLASPIILAPVAAHKLANEQGEIASAK 116

Query: 123 GLKEFGSIYTTSSYSTTDLPEISAALGGTPHWFQFYYSKDDGINRNIMDRVKAQGCKAIV 182
            G+KEFG+IYTTSSYSTTDLPEIS  LG +PHWFQFYYSKDDGINR+IMDR+KA+G K+IV
Sbjct: 117 GVKEFGTIYTTSSYSTTDLPEISQTLGDSPHWFQFYYSKDDGINRHIMDRLKAEGVKSIV 176

Query: 183 LTADATVGGNREVDRRNGFVFPVGMPIVQEYLPDGAGKTMDYVYKSAKQALTSKDIEYIA 242
            LT DATVGGNREVD+RNGFVFPVGMPIVQEYLP+GAGKTMDYVYK+ KQAL+ KD+EYIA
Sbjct: 177 LTVDATVGGNREVDKRNGFVFPVGMPIVQEYLPNGAGKTMDYVYKATKQALSPKDVEYIA 236

Query: 243 TYSGLPVYVKGPQCAEDTLRALDAGASGIWVTNHGGRQLDGGPAAFDSLQEVAEAVDQKV 302
             YSGLPVYVKGPQCAED  RAL+AGASGIWVTNHGGRQLDGGPAAFDSLQEVAE+VD++V
Sbjct: 237 QYSGLPVYVKGPQCAEDAFRALEAGASGIWVTNHGGRQLDGGPAAFDSLQEVAESVDRRV 296

Query: 303 PIVFDSGIRRGQHIFKALASGADLVALGRPAIYGLAMGGSIGTRQVFEKLNDELKMVMQL 362
            PIVFDSG+RRGQH+FKALASGADLVALGRP IYGLAMGGS+GTRQVFEK+NDELKMVMQL
Sbjct: 297 PIVFDSGVRRGQHVFKALASGADLVALGRPVIYGLAMGGSVGTRQVFEKINDELKMVMQL 356

Query: 363 AGTQTIQDVKAFNLRHNPYDSSIPFDQNALRL 394
            AGTQTI DVK F LRHNPYDSSIPF     ++
Sbjct: 357 AGTQTIDDVKHFKLRHNPYDSSIPFSPKCFKI 388
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2470**

A DNA sequence (GASx414) was identified in *S.pyogenes* <SEQ ID 7439> which encodes the amino acid sequence <SEQ ID 7440>. Analysis of this protein sequence reveals the following:

```
Possible site: 32

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0682(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2471**

A DNA sequence (GASx417R) was identified in *S.pyogenes* <SEQ ID 7441> which encodes the amino acid sequence <SEQ ID 7442>. Analysis of this protein sequence reveals the following:

```
Possible site: 34

>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1765(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2472**

A DNA sequence (GASx418) was identified in *S.pyogenes* <SEQ ID 7443> which encodes the amino acid sequence <SEQ ID 7444>. Analysis of this protein sequence reveals the following:

```
Possible site: 32

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.2532(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2473**

A DNA sequence (GASx419) was identified in *S.pyogenes* <SEQ ID 7445> which encodes the amino acid sequence <SEQ ID 7446>. Analysis of this protein sequence reveals the following:

```
Possible site: 28

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.3082(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2474**

A DNA sequence (GASx423) was identified in *S.pyogenes* <SEQ ID 7447> which encodes the amino acid sequence <SEQ ID 7448>. Analysis of this protein sequence reveals the following:

```
Possible site: 52
```

```
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -2.18    Transmembrane   14 -  30 (  13 -  31)

----- Final Results -----
               bacterial membrane --- Certainty=0.1871(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2475**

A DNA sequence (GASx427R) was identified in *S.pyogenes* <SEQ ID 7449> which encodes the amino acid sequence <SEQ ID 7450>. Analysis of this protein sequence reveals the following:

```
Possible site: 25

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -1.17    Transmembrane   13 -  29 (  10 -  29)

----- Final Results -----
               bacterial membrane --- Certainty=0.1468(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related sequence was also identified in GAS <SEQ ID 9105> which encodes the amino acid sequence <SEQ ID 9106>. Analysis of this protein sequence reveals the following:

```
     Possible site: 20
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -1.17    Transmembrane    8 -  24

----- Final Results -----
               bacterial membrane --- Certainty=0.1470(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAA26616 GB:M63917 epidermal cell differentiation inhibitor
          [Staphylococcus aureus]
 Identities = 58/195 (29%), Positives = 106/195 (53%), Gaps = 13/195 (6%)

Query: 67  RWGKGLI----YPRAEQEAMAAYTCQQAGPINTSLDKAKGELSQLTPELRDQVAQLDAAT 122
            +WG  LI    Y   ++ A+  YT + +  IN  L  A G++++L    +D+V +LD++
Sbjct: 49  KWGNKLIKQAKYSSDDKIALYEYT-KDSSKINGPLRLAGGDINKLDSTTQDKVRRLDSSI 107

Query: 123 HRLVIPWNIVVYRYVYETFLRDI-GVSHADLTSYYR--NHQFDPHILCKIK--LGTR-YT 176
             +   P ++ VYR +   +L  I G ++ DL    +  N Q+D +++ K+   + +R Y
Sbjct: 108 SKSTTPESVYVYRLLNLDYLTSIVGFTNEDLYKLQQTNNGQYDENLVRKLNNVMNSRIYR 167

Query: 177 KHSFMSTTALKNGAMTHRPVEVRICVKKGAKAAFV--EPYSAVPSEVELLFPRGCQLEVV 234
            +  + ST  +   A+  RP+E+R+ + KG KAA++  +  +A   + E+L PRG +  V
Sbjct: 168 EDGYSSTQLVSGAAVGGRPIELRLELPKGTKAAYLNSKDLTAYYGQQEVLLPRGTEYAVG 227

Query: 235 GAYVSQDQKKLHIEA 249
               +S D+KK+ I A
```

```
Sbjct: 228 SVELSNDKKKIIITA 242
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2476**

A DNA sequence (GASx428) was identified in *S.pyogenes* <SEQ ID 7451> which encodes the amino acid sequence <SEQ ID 7452>. Analysis of this protein sequence reveals the following:

```
Possible site: 14

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3817(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2477**

A DNA sequence (GASx429) was identified in *S.pyogenes* <SEQ ID 7453> which encodes the amino acid sequence <SEQ ID 7454>. Analysis of this protein sequence reveals the following:

```
Possible site: 32

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2478**

A DNA sequence (GASx431) was identified in *S.pyogenes* <SEQ ID 7455> which encodes the amino acid sequence <SEQ ID 7456>. Analysis of this protein sequence reveals the following:

```
Possible site: 43

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -8.60   Transmembrane   68 -  84 (  66 -  90)
   INTEGRAL    Likelihood = -6.85   Transmembrane   22 -  38 (  16 -  42)
   INTEGRAL    Likelihood = -3.29   Transmembrane   44 -  60 (  43 -  61)

----- Final Results -----
               bacterial membrane --- Certainty=0.4439(Affirmative) < succ>
```

```
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2479**

A DNA sequence (GASx432R) was identified in *S.pyogenes* <SEQ ID 7457> which encodes the amino acid sequence <SEQ ID 7458>. Analysis of this protein sequence reveals the following:

```
Possible site: 22

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2480**

A DNA sequence (GASx434) was identified in *S.pyogenes* <SEQ ID 7459> which encodes the amino acid sequence <SEQ ID 7460>. Analysis of this protein sequence reveals the following:

```
Possible site: 24

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2481**

A DNA sequence (GASx435R) was identified in *S.pyogenes* <SEQ ID 7461> which encodes the amino acid sequence <SEQ ID 7462>. Analysis of this protein sequence reveals the following:

```
Possible site: 25

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -2.50   Transmembrane    4 -  20 (   3 -  21)
```

```
----- Final Results -----
               bacterial membrane --- Certainty=0.1999(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB59092 GB:M97157 pyrogenic exotoxin C [Streptococcus pyogenes]
 Identities = 110/229 (48%), Positives = 150/229 (65%), Gaps = 4/229 (1%)

Query: 4   IIKTIILVIIIFHGYGS--VKSDSE-NIKDVKLQLNYAYEIIPVDYTNCNIDYLTTHDFY 60
           IIK + ++ +I     S  +KSDS+ +I +VK  L YAY I P DY +C +++ TTH
Sbjct: 6   IIKIVFIITVILISTISPIIKSDSKKDISNVKSDLLYAYTITPYDYKDCRVNFSTTHTLN 65

Query: 61  IDISSYKKKNFSVDSEVESYITTKFTKNQKVNIFGLPYIFTRYDVYYIYGGVTPSVNSNS 120
           ID   Y+ K++ + SE+    + KF ++  V++FGL YI   +   YIYGG+TP+ N N
Sbjct: 66  IDTQKYRGKDYYISSEMSYEASQKFKRDDHVDVFGLFYILNSHTGEYIYGGITPAQN-NK 124

Query: 121 ENSKIVGNLLIDGVQQKTLINPIKIDKPIFTIQEFDFKIRQYLMQTYKIYDPNSPYIKGQ 180
            N K++GNL I G  Q+ L N I ++K I T QE DFKIR+YLM  YKIYD  SPY+ G+
Sbjct: 125 VNHKLLGNLFISGESQQNLNNKIILEKDIVTFQEIDFKIRKYLMDNYKIYDATSPYVSGR 184

Query: 181 LEIAINGNKHESFNLYDATSSSTRSDIFKKYKDNKTINMKDFSHFDIYL 229
           +EI     KHE  +L+D+ +  TRSDIF KYKDN+ INMK+FSHFDIYL
Sbjct: 185 IEIGTKDGKHEQIDLFDSPNEGTRSDIFAKYKDNRIINMKNFSHFDIYL 233
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2482**

A DNA sequence (GASx436R) was identified in *S.pyogenes* <SEQ ID 7463> which encodes the amino acid sequence <SEQ ID 7464>. Analysis of this protein sequence reveals the following:

```
Possible site: 22

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2483**

A DNA sequence (GASx446) was identified in *S.pyogenes* <SEQ ID 7465> which encodes the amino acid sequence <SEQ ID 7466>. Analysis of this protein sequence reveals the following:

```
Possible site: 20

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
```

```
                   bacterial outside --- Certainty=0.3000(Affirmative) < succ>
                  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2484**

A DNA sequence (GASx449) was identified in *S.pyogenes* <SEQ ID 7467> which encodes the amino acid sequence <SEQ ID 7468>. Analysis of this protein sequence reveals the following:

```
Possible site: 15

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -3.82    Transmembrane    3 -  19 (   1 -  20)

----- Final Results -----
              bacterial membrane --- Certainty=0.2529(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2485**

A DNA sequence (GASx450R) was identified in *S.pyogenes* <SEQ ID 7469> which encodes the amino acid sequence <SEQ ID 7470>. Analysis of this protein sequence reveals the following:

```
Possible site: 30

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -1.44    Transmembrane   21 -  37 (  19 -  37)

----- Final Results -----
              bacterial membrane --- Certainty=0.1574(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2486**

A DNA sequence (GASx457R) was identified in *S.pyogenes* <SEQ ID 7471> which encodes the amino acid sequence <SEQ ID 7472>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
```

```
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-15.34    Transmembrane    64 -  80 (  57 -  86)
   INTEGRAL    Likelihood =-13.43    Transmembrane    97 - 113 (  91 - 116)
   INTEGRAL    Likelihood = -5.57    Transmembrane    38 -  54 (  32 -  56)

----- Final Results -----
              bacterial membrane --- Certainty=0.7135(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2487

A DNA sequence (GASx476R) was identified in *S.pyogenes* <SEQ ID 7473> which encodes the amino acid sequence <SEQ ID 7474>. Analysis of this protein sequence reveals the following:

```
Possible site: 31

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.3013(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2488

A DNA sequence (GASx477) was identified in *S.pyogenes* <SEQ ID 7475> which encodes the amino acid sequence <SEQ ID 7476>. Analysis of this protein sequence reveals the following:

```
Possible site: 57

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1022(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAC03521 GB:AJ276410 BlpJ protein [Streptococcus pneumoniae]
 Identities = 47/77 (61%), Positives = 59/77 (76%)

Query: 1  MIKFAEEIQKEELFHIIGGYSATDCKNHLIGGITSGAIAGGVGAGMATLGVGGVAGAFAG 60
          M+   E + E L  + GGYS+TDC+N LI G+T+G I GG GAG+ATLGV G+AGAF G
Sbjct: 5  MLSQLEVMDTEMLAKVEGGYSSTDCQNALITGVTTGIITGGTGAGLATLGVAGLAGAFVG 64
```

```
Query: 61 AHVGAIAGGLTCVGGML 77
          AH+GAI GGLTC+GGM+
Sbjct: 65 AHIGAIGGGLTCLGGMV 81
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2489**

A DNA sequence (GASx478) was identified in *S.pyogenes* <SEQ ID 7477> which encodes the amino acid sequence <SEQ ID 7478>. Analysis of this protein sequence reveals the following:

```
Possible site: 45

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.07    Transmembrane   42 -  58 (  41 -  58)
   INTEGRAL    Likelihood = -1.59    Transmembrane   22 -  38 (  22 -  38)

----- Final Results -----
               bacterial membrane --- Certainty=0.1829(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAC03520 GB:AJ276410 BlpI protein [Streptococcus pneumoniae]
 Identities = 35/56 (62%), Positives = 44/56 (78%)

Query: 1  MDNFLELQFEELVNISGGKGNIGSAIGGCLGGMLIAAAGGPITGGAAAFVCVASGI 56
          M+ F  +  EEL  +SGG+GN+GSAIGGC+G +L+AAA GPITGGAA  +CV SGI
Sbjct: 6  MEQFSVMDNEELEIVSGGRGNLGSAIGGCIGAVLLAAATGPITGGAATLICVGSGI 61
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2490**

A DNA sequence (GASx482) was identified in *S.pyogenes* <SEQ ID 7479> which encodes the amino acid sequence <SEQ ID 7480>. Analysis of this protein sequence reveals the following:

```
Possible site: 14

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -0.43    Transmembrane   61 -  77 (  61 -  79)

----- Final Results -----
               bacterial membrane --- Certainty=0.1171(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAC03524 GB:AJ276410 BlpM protein [Streptococcus pneumoniae]
 Identities = 22/52 (42%), Positives = 30/52 (57%)

Query: 29 MEIKKLETFHQMTIEKLAKVEGGKNNWQANVSGVIAAGSAGAAIGFPVCGVA 80
          M+ K +E FH+M I  L+ +EGGKNNWQ NV     A   G  +G  +C  +
Sbjct: 1  MDTKIMEQFHEMDITMLSSIEGGKNNWQTNVLEGGGAAFGGWGLGTAICAAS 52
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2491**

A DNA sequence (GASx483) was identified in *S.pyogenes* <SEQ ID 7481> which encodes the amino acid sequence <SEQ ID 7482>. Analysis of this protein sequence reveals the following:

```
Possible site: 58

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1832(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2492**

A DNA sequence (GASx484) was identified in *S.pyogenes* <SEQ ID 7483> which encodes the amino acid sequence <SEQ ID 7484>. Analysis of this protein sequence reveals the following:

```
Possible site: 21

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2493**

A DNA sequence (GASx485) was identified in *S.pyogenes* <SEQ ID 7485> which encodes the amino acid sequence <SEQ ID 7486>. Analysis of this protein sequence reveals the following:

```
Possible site: 32

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1037(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2494**

A DNA sequence (GASx487) was identified in *S.pyogenes* <SEQ ID 7487> which encodes the amino acid sequence <SEQ ID 7488>. Analysis of this protein sequence reveals the following:

```
Possible site: 50

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1086(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2495**

A DNA sequence (GASx488) was identified in *S.pyogenes* <SEQ ID 7489> which encodes the amino acid sequence <SEQ ID 7490>. Analysis of this protein sequence reveals the following:

```
Possible site: 22

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2176(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2496**

A DNA sequence (GASx489R) was identified in *S.pyogenes* <SEQ ID 7491> which encodes the amino acid sequence <SEQ ID 7492>. Analysis of this protein sequence reveals the following:

```
Possible site: 22

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2497**

A DNA sequence (GASx490) was identified in *S.pyogenes* <SEQ ID 7493> which encodes the amino acid sequence <SEQ ID 7494>. Analysis of this protein sequence reveals the following:

```
Possible site: 24

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.2547(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2498**

A DNA sequence (GASx491R) was identified in *S.pyogenes* <SEQ ID 7495> which encodes the amino acid sequence <SEQ ID 7496>. Analysis of this protein sequence reveals the following:

```
Possible site: 22

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-10.24   Transmembrane    6 -  22 (   3 -  28)

----- Final Results -----
              bacterial membrane --- Certainty=0.5097(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2499**

A DNA sequence (GASx492) was identified in *S.pyogenes* <SEQ ID 7497> which encodes the amino acid sequence <SEQ ID 7498>. Analysis of this protein sequence reveals the following:

```
Possible site: 27

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```

```
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2500**

A DNA sequence (GASx493) was identified in *S.pyogenes* <SEQ ID 7499> which encodes the amino acid sequence <SEQ ID 7500>. Analysis of this protein sequence reveals the following:

```
Possible site: 19

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.69   Transmembrane   21 -  37 (  21 -  37)

----- Final Results -----
                  bacterial membrane --- Certainty=0.1277(Affirmative) < succ>
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2501**

A DNA sequence (GASx495R) was identified in *S.pyogenes* <SEQ ID 7501> which encodes the amino acid sequence <SEQ ID 7502>. Analysis of this protein sequence reveals the following:

```
Possible site: 28

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
                 bacterial cytoplasm --- Certainty=0.2891(Affirmative) < succ>
                  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2502**

A DNA sequence (GASx499R) was identified in *S.pyogenes* <SEQ ID 7503> which encodes the amino acid sequence <SEQ ID 7504>. Analysis of this protein sequence reveals the following:

```
Possible site: 15

>>> Seems to have an uncleavable N-term signal seq
```

```
   INTEGRAL    Likelihood = -2.50   Transmembrane    3 -  19 (   1 -  20)

----- Final Results -----
               bacterial membrane --- Certainty=0.1999(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2503**

A DNA sequence (GASx500) was identified in *S.pyogenes* <SEQ ID 7505> which encodes the amino acid sequence <SEQ ID 7506>. Analysis of this protein sequence reveals the following:

```
Possible site: 54

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC77220 GB:AE000497 orf, hypothetical protein [Escherichia
          coli]
 Identities = 262/480 (54%), Positives = 338/480 (69%), Gaps = 5/480 (1%)

Query: 18  GMLNRHGLIAGATGTGKTVTLKVLAEQLSLAGVPVFLADIKGDLSNLTKAGEVTDKLAAR 77
           GM NRHGLI GATGTGKTVTL+ LAE LS  GVPVF+AD+KGDL+ + +AG V++KL AR
Sbjct: 20  GMANRHGLITGATGTGKTVTLQKLAESLSEIGVPVFMADVKGDLTGVAQAGTVSEKLLAR 79

Query: 78  LATIGVSDYQPQAFPVRMWDVFGQNGQPLRTTISELGPMMLSRLLNLNDTQTGVLNIVFK 137
           L  IGV+D+QP A PV +WD+FG+ G P+R T+S+LGP++L+RLLNLND Q+GVLNI+F+
Sbjct: 80  LKNIGVNDWQPHANPVVVWDIFGEKGHPVRATVSDLGPLLLARLLNLNDVQSGVLNIIFR 139

Query: 138 IADEKGWLLIDLKDLQAILKEVGDHASDYSSHYGNIAKQSIGAIQRSLLTLEQEGAHQFF 197
           IAD++G LL+D KDL+AI + +GD+A  + + YGNI+  S+GAIQR LL+LEQ+GA  FF
Sbjct: 140 IADDQGLLLLDFKDLRAITQYIGDNAKSFQNQYGNISSASVGAIQRGLLSLEQQGAAHFF 199

Query: 198 GEPALDVADLMQLDVASGYGAINILSATKLFQSPTLYTTFLLWLLSELYKLLPEVGDLDK 257
           GEP LD+ D M+ D A+G G INILSA KL+Q P LY   LLW+LSELY+ LPE GDL+K
Sbjct: 200 GEPMLDIKDWMRTD-ANGKGVINILSAEKLYQMPKLYAASLLWMLSELYEQLPEAGDLEK 258

Query: 258 PKMVFFFDEAHLLFKDAPKVFLEKVEQIVRLIRSKGVGIFFVTQNPLDLPETVLAQLGNR 317
           PK+VFFFDEAHLLF DAP+V L+K+EQ++RLIRSKGVG++FV+QNP D+P+ VL QLGNR
Sbjct: 259 PKLVFFFDEAHLLFNDAPQVLLDKIEQVIRLIRSKGVGVWFVSQNPSDIPDNVLGQLGNR 318

Query: 318 IQHAFRAYTPKEQKAVRVAADTFRQNPDLDVARVITELEVGEALISVLNDKGQPSIVERA 377
           +QHA RA+TPK+QKAV+ AA T R NP  D  + I EL  GEALIS L+ KG PS+VERA
Sbjct: 319 VQHALRAFTPKDQKAVKAAAQTMRANPAFDTEKAIQELGTGEALISFLDAKGSPSVVERA 378

Query: 378 YIMPPKSSFAVLSEIESQQLVQSSPFASKYSQSIDRESAYEKLAAKVLEDNRLAQEAIAT 437
            ++ P S    ++E E   L+  SP   KY   +DRESAYE L  K  + +   Q
Sbjct: 379 MVIAPCSRMGPVTEDERNGLINHSPVYGKYEDEVDRESAYEML-QKGFQASTEQQNNPPA 437

Query: 438 AQREKEAKEAIKAQAATKKANRRSVGRSHKTVVEKATDAFISTTVRTIGRELVRGLLGSL 497
```

```
                +E    + I                    K    +         +  R +  ++VRG+LGSL
Sbjct: 438 KGKEVAVDDGILGGLKDILFGTTGPRGGKK---DGVVQTMAKSAARQVTNQIVRGMLGSL 494
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2504**

A DNA sequence (GASx502) was identified in *S.pyogenes* <SEQ ID 7507> which encodes the amino acid sequence <SEQ ID 7508>. Analysis of this protein sequence reveals the following:

```
Possible site: 49

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-13.59   Transmembrane   59 -  75 (  52 -  77)
   INTEGRAL    Likelihood = -9.34   Transmembrane    4 -  20 (   1 -  24)

----- Final Results -----
              bacterial membrane --- Certainty=0.6434(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB15368 GB:Z99121 yvaL [Bacillus subtilis]
 Identities = 28/72 (38%), Positives = 44/72 (60%), Gaps = 2/72 (2%)

Query: 1  MYNLLLTILLVLSGLLEIAIFMQPQKNPSSNVFDSSGSEALFERTKARGFEAFMQRFTAV 60
           M+ +L+T+L+++S  L I + +Q  K+   +   S G+E LF + KARG +  + R T V
Sbjct: 1  MHAVLITLLVIVSIALIIVVLLQSSKSAGLSGAISGGAEQLFGKQKARGLDLILHRITVV 60

Query: 61 L--VFFWLAIAL 70
           L  +FF L IAL
Sbjct: 61 LAVLFFVLTIAL 72
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2505**

A DNA sequence (GASx505) was identified in *S.pyogenes* <SEQ ID 7509> which encodes the amino acid sequence <SEQ ID 7510>. Analysis of this protein sequence reveals the following:

```
Possible site: 45

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.44   Transmembrane  140 - 156 ( 138 - 156)

----- Final Results -----
              bacterial membrane --- Certainty=0.1574(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF09704 GB:AE001874 glutamine cyclotransferase [Deinococcus radiodurans]
 Identities = 81/229 (35%), Positives = 128/229 (55%), Gaps = 10/229 (4%)

Query: 16  YSYDSNLYTQGLEQLNNNHILLSAGRYGFSKVGVYDL--TQEIFSEKIAFP-DTVFAEGL 72
```

```
           Y +D   +TQGL+ L   H L S G+ G S + V +L   + ++S  +A      F EG
Sbjct: 54  YPHDRAAFTQGLQYLGGGHYLESTGQVGESDLRVSELRGAKVLWSTPLAQALPQAFGEGS 113

Query: 73  TVVEDYFWLLTYKEGVAYKFDKATCNCLGAYPFEGDGWGLAYDKENQCLWMTSGNAFLQK 132
           T +    + LT+++GVA  +D  T    G + ++G+GWGL  D ++  L M++G + L
Sbjct: 114 TQLGSTVYQLTWQDGVALTYDARTFKETGRHRYQGEGWGLTSDGKS--LIMSNGTSTLVW 171

Query: 133 RDPKDFALLDTVLVAIESVPISMLNELEYVDGYLYANIWQTNTIVKLQPDSGKVVATYDI 192
           RDPK FA   +V V  +  P+  LNELEYV G +YAN+W T+ I ++ P +GKV+   D+
Sbjct: 172 RDPKTFAAQRSVQVTDQGQPVRNLNELEYVQGSVYANVWLTDRIARIHPQTGKVLTWIDV 231

Query: 193 SPLLKALNLDKSHYPDL----NVLNGIAHLDQQ-RFLITGKLYPLMLEV 236
           S L + ++    +        +V NGIA + ++  L+TGK +P + EV
Sbjct: 232 SDLTREVSAAATKQGQALTFDDVPNGIAFIPERGTLLLTGKRWPTLFEV 280
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2506**

A DNA sequence (GASx506R) was identified in *S.pyogenes* <SEQ ID 7511> which encodes the amino acid sequence <SEQ ID 7512>. Analysis of this protein sequence reveals the following:

```
Possible site: 33

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2800(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2507**

A DNA sequence (GASx507R) was identified in *S.pyogenes* <SEQ ID 7513> which encodes the amino acid sequence <SEQ ID 7514>. Analysis of this protein sequence reveals the following:

```
Possible site: 53

>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-10.51   Transmembrane  103 - 119 (  97 - 124)
   INTEGRAL    Likelihood = -9.13   Transmembrane  126 - 142 ( 122 - 145)
   INTEGRAL    Likelihood = -8.65   Transmembrane  290 - 306 ( 286 - 307)
   INTEGRAL    Likelihood = -7.17   Transmembrane  200 - 216 ( 198 - 228)
   INTEGRAL    Likelihood = -7.06   Transmembrane   58 -  74 (  54 -  82)
   INTEGRAL    Likelihood = -3.19   Transmembrane  223 - 239 ( 220 - 242)
   INTEGRAL    Likelihood = -2.81   Transmembrane  244 - 260 ( 244 - 261)
   INTEGRAL    Likelihood = -2.71   Transmembrane  174 - 190 ( 169 - 191)

----- Final Results -----
               bacterial membrane --- Certainty=0.5203(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB56669 GB:AL121596 putative membrane protein [Streptomyces
          coelicolor A3(2)]
 Identities = 119/322 (36%), Positives = 182/322 (55%), Gaps = 24/322 (7%)

Query: 9   LETIYILIGLQLFHTAYCTFKDKTNPVYFGTALFWGLLGVTFV------------GGAFL 56
           +E +Y LIGL     A     D++NP  + +A FWGLLG+TF             GG  L
Sbjct: 4   VEWLYWLIGLVFVVMAVQMAMDRSNPKRWTSAAFWGLLGLTFPYGTGVANATAGNGGWTL 63

Query: 57  PNKVIGFIVIVLALLTLFKQVRIGTLPAFNEQKAEESAHRIGNWIFLPVMLMAMISLLLA 116
           P + +G  V+ L +L  F  +  G       ++ E +A R+GN IF+P + + +++++ A
Sbjct: 64  PAEPLGVAVLALIVLAGFNFLGKGVPVTTTGEQREAAAARLGNKIFVPALTIPLVAIVCA 123

Query: 117 LILPDFSKSAIGIAGILA---------TIAILIITKQKPSALLAENNRMNQQVSTSGILP 167
            +L +      G A +L           + +L+  ++K S  +     M + + ++ +LP
Sbjct: 124 SVLDESGLFETGKATLLGLGLGCVAALVVGMLVTGEKKLSVPIHSGRSMLEAMGSALLLP 183

Query: 168 QLLGALGAIFAAAGVGDVIASLIREIVPADSRFFGVLAYVLGMVIFTMIMGNAFAAFTVI 227
           QLL  LG+IFAAAGVGD +  ++ +++P DS++F VLAY +GM +FT+IMGNAFAAF V+
Sbjct: 184 QLLAVLGSIFAAAGVGDQVGDIMNKVLPDDSKYFAVLAYCVGMFLFTVIMGNAFAAFPVM 243

Query: 228 TTGIGVPFVFAL--GADPIIAGALAMTAGFCGTLLTPMAANFNALPVALMEIKDRNAVIK 285
           T  IG P +     G +P +  A+ M AGF GTL TPMAANFN +P  L+E+KD+   IK
Sbjct: 244 TAAIGWPVLIQQMHGNEPAVL-AIGMLAGFAGTLCTPMAANFNIVPATLLELKDQYGPIK 302

Query: 286 KQAPIALVLIISHIALMYLLAY 307
            Q P  + L+     +M L A+
Sbjct: 303 AQLPTGIALLGCCTVIMALFAF 324
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2508**

A DNA sequence (GASx508R) was identified in *S.pyogenes* <SEQ ID 7515> which encodes the amino acid sequence <SEQ ID 7516>. Analysis of this protein sequence reveals the following:

```
Possible site: 61

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-12.15    Transmembrane  212 - 228 ( 208 - 235)
   INTEGRAL    Likelihood = -8.81    Transmembrane   23 -  39 (  17 -  64)
   INTEGRAL    Likelihood = -7.43    Transmembrane   45 -  61 (  40 -  64)
   INTEGRAL    Likelihood = -1.49    Transmembrane  114 - 130 ( 114 - 130)
   INTEGRAL    Likelihood = -1.49    Transmembrane    3 -  19 (   3 -  20)
   INTEGRAL    Likelihood = -1.49    Transmembrane   76 -  92 (  76 -  92)

----- Final Results -----
               bacterial membrane --- Certainty=0.5861(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB56670 GB:AL121596 possible integral membrane protein
          [Streptomyces coelicolor A3(2)]
 Identities = 77/220 (35%), Positives = 138/220 (62%), Gaps = 2/220 (0%)

Query: 23  IKLIGIVIIVLGFILKCDAIATVVVAGLVTALVSGISFIDFLDILGKEFTNQRLLTIFFI 82
           I L+G+V+++LGF+ + + +  V VAG+VT L+  ++ ++ L   G+ F + R +T++ I
Sbjct: 2   IVLLGVVVVILGFVTRRNPVLVVGVAGIVTGLLGKMNPLEVLAAFGRSFADSRSVTVYAI 61

Query: 83  TLPLIGLSETYGLKHRATQLIQRVQALTVGRLLTLYLIIRELAGLFSIR-LGGHPQFVRP 141
            LP+IGL E YGL+ +A  LI R+  L+ GR LT+YL++R++   F +  +GG  Q VRP
Sbjct: 62  VLPVIGLLERYGLREQARHLIGRLGKLSAGRFLTVYLLVRQVTAAFGLNSIGGPAQTVRP 121
```

```
Query: 142 LIQPMGEAAAKANIGEELTDAEKDDIKAMAAANENFGNFFAQNTFVGAGGVLLIAGTLEQ 201
           L+ PM EAAA+ + G +L D  ++ +++ +A+ +  G FF ++ F+  G +LLI G +
Sbjct: 122 LVAPMAEAAAERSTGAKLPDKLREKVRSYSASADTVGVFFGEDCFIAIGSILLITGFVNS 181

Query: 202 LGY-DGNQAKIAFSSILIAIISIIIVAIYNYLFEKKMERQ 240
             + D    ++A  +I +A+ + +I      L +K++ER+
Sbjct: 182 TYHQDIEPTQLALWAIPLAVCAFLIHGARLLLMDKQLERE 221
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2509**

A DNA sequence (GASx520) was identified in *S.pyogenes* <SEQ ID 7517> which encodes the amino acid sequence <SEQ ID 7518>. Analysis of this protein sequence reveals the following:

```
Possible site: 13

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2652(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2510**

A DNA sequence (GASx522R) was identified in *S.pyogenes* <SEQ ID 7519> which encodes the amino acid sequence <SEQ ID 7520>. Analysis of this protein sequence reveals the following:

```
Possible site: 21

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2511**

A DNA sequence (GASx523) was identified in *S.pyogenes* <SEQ ID 7521> which encodes the amino acid sequence <SEQ ID 7522>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
```

```
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2133(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2512**

A DNA sequence (GASx525) was identified in *S.pyogenes* <SEQ ID 7523> which encodes the amino acid sequence <SEQ ID 7524>. Analysis of this protein sequence reveals the following:

```
Possible site: 14

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2364(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2513**

A DNA sequence (GASx535) was identified in *S.pyogenes* <SEQ ID 7525> which encodes the amino acid sequence <SEQ ID 7526>. Analysis of this protein sequence reveals the following:

```
Possible site: 47

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4223(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2514**

A DNA sequence (GASx536) was identified in *S.pyogenes* <SEQ ID 7527> which encodes the amino acid sequence <SEQ ID 7528>. Analysis of this protein sequence reveals the following:

```
Possible site: 59

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1102(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB85515 GB:AE000874 conserved protein [Methanobacterium
           thermoautotrophicum]
 Identities = 82/236 (34%), Positives = 132/236 (55%), Gaps = 11/236 (4%)

Query: 9   MNLSIFGLKNIPYLKEGDSIEKLIEESIKTSEFFIEDNDVLCIASKVVSIAEGQVMSLNE 68
           M +S+ G++ +P +  GD I  LI  ++      + D D++ IA  +VS AEG ++SL E
Sbjct: 1   MGISLIGVEGMPLVGAGDDIAYLIISALNEGGEDLLDGDIIVIAETIVSKAEGNIISLEE 60

Query: 69  IQVSDVAKEIHRNIPRKDPRIIEIMLNLVNRDLSRLDIKKNYIGCRLENGLKLTSGGIDR 128
           I+ S  A +I      KDP ++E +L         + +  ++I     +G    + GID
Sbjct: 61  IKPSPEALDIAERTG-KDPSLVEAILG---ESSEIIRVGHDFIVSETRHGFVCANAGIDE 116

Query: 129 KSVDEVFL--LPNNPDASAKRISEYLKKSLGKNVAVVITDSDGREDKRGATQVAIGIYGI 186
            +VD+     LP +PD SA++I   L+++ G+ +AV+I+D+ GR  + GA  VA+G+ G+
Sbjct: 117 SNVDDGLATPLPRDPDGSAEKILRTLQEATGRELAVIISDTQGRPFREGAVGVAVGVAGL 176

Query: 187 HPL--RKTEVIDSQGETIKFQEETLCDMIAACAGLVMGQRGTGIPAVLIRGLDYKW 240
            P+  RK E  D  G +++     + D +AA A LVMGQ   G+PAV+IRG  Y W
Sbjct: 177 SPIWDRKGE-RDLYGRSLETTRVAVADELAAAASLVMGQADEGVPAVIIRG--YPW 229
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2515**

A DNA sequence (GASx537) was identified in *S.pyogenes* <SEQ ID 7529> which encodes the amino acid sequence <SEQ ID 7530>. Analysis of this protein sequence reveals the following:

```
Possible site: 50

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.12   Transmembrane  174 - 190 ( 174 - 190)

----- Final Results -----
               bacterial membrane --- Certainty=0.1447(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2516**

A DNA sequence (GASx538) was identified in *S.pyogenes* <SEQ ID 7531> which encodes the amino acid sequence <SEQ ID 7532>. Analysis of this protein sequence reveals the following:

```
Possible site: 32

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3852(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB99212 GB:U67562 conserved hypothetical protein [Methanococcus
           jannaschii]
 Identities = 129/387 (33%), Positives = 208/387 (53%), Gaps = 44/387 (11%)

Query: 18  EVVERKGLGHPDTLADGIAEQIEIDYSLYCLDKFGVIPHHNFDKIIIRGGHSVQDFGGSD 77
           E+VERKGLGHPD++ DGIAE +        ++KFG I HHN D++ + GGH+   FGG
Sbjct: 20  EIVERKGLGHPDSICDGIAESVSRALCKMYMEKFGTILHHNTDQVELVGGHAYPKFGGGV 79

Query: 78  FIEPIKIIFLGRASKKCFNS------SIPLFKIQKKAATKYLNRILPNLDVENYVEFETL 131
            + PI I+  GRA+ +  +        +P+     KAA +YL ++L N+DV+  V  +
Sbjct: 80  MVSPIYILLSGRATMEILDKEKNEVIKLPVGTTAVKAAKEYLKKVLRNVDVDKDVIID-- 137

Query: 132 TSDFTTKTNWFSPEAIEDLP-EYLDVPKANDTATMISYWPLTISEELALMIEGYFYKLD- 189
                 +    S + ++    +  +VP ANDT+  + Y PL+ +E L L  E +    +
Sbjct: 138 -----CRIGQGSMDLVDVFERQKNEVPLANDTSFGVGYAPLSTTERLVLETERFLNSDEL 192

Query: 190 KNELPTPRFTKMGGDIKVMVVRNDLEYSIRINFPLISKFFNNDIESQLYVDKHVEKIKKY 249
           KNE+P      +G DIKVM +R   + ++ I   ++ ++  N IE    V   +EK++K
Sbjct: 193 KNEIPA-----VGEDIKVMGLREGKKITLTIAMAVVDRYVKN-IEEYKEV---IEKVRKK 243

Query: 250 IEQKYKNIS--FSIDYH-----------YYLTTTGSCIDFGEEGAVGRGNKTHGIISSFR 296
           +E   K I+  + ++ H            YLT TG+  + G++G+VGRGN+ +G+I+ FR
Sbjct: 244 VEDLAKKIADGYEVEIHINTADDYERESVYLTVTGTSAEMGDDGSVGRGNRVNGLITPFR 303

Query: 297 PNTMEAPAGKNCTYFVGKVWGFLSDTIAKEIYEAFNT-PCQIIMQLNIGSKLYRPTHLFI 355
           P +MEA +GKN    VGK++  L++ IA +I +      C + +   IG  +  P  L I
Sbjct: 304 PMSMEAASGKNPVNHVGKIYNILANLIANDIAKLEGVKECYVRILSQIGKPINEPKALDI 363

Query: 356 Q--TEESVD----QERVLEIVNRHLNN 376
           +  TE+S D    + +  EI N+ L+N
Sbjct: 364 EIITEDSYDIKDIEPKAKEIANKWLDN 390
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2517**

A DNA sequence (GASx539) was identified in *S.pyogenes* <SEQ ID 7533> which encodes the amino acid sequence <SEQ ID 7534>. Analysis of this protein sequence reveals the following:

```
Possible site: 17

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.1436(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2518**

A DNA sequence (GASx540) was identified in *S.pyogenes* <SEQ ID 7535> which encodes the amino acid sequence <SEQ ID 7536>. Analysis of this protein sequence reveals the following:

```
Possible site: 45

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3956(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAD36304 GB:AE001779 conserved hypothetical protein [Thermotoga maritima]
 Identities = 105/353 (29%), Positives = 173/353 (48%), Gaps = 46/353 (13%)

Query: 3   VIGIPTLNEADNISRLVKQIDEYAVNL-GKEIIIINSDSKSTDGTPQIFLETKTYNT-KV 60
           V+GIP+ N A+ IS + +   +  V+    + +I+NSD  S DGT + F+ET T+   K
Sbjct: 106 VVGIPSYNNAETISHVARTAAQGIVDFFDGDGMIVNSDGGSADGTRERFMETDTFGLPKE 165

Query: 61  SIVSEA-KGKGYNVRNIFEYAINHVPNFSGLILIDGDVVSMKKMWLEKMFIAIESGN-DL 118
           S V E   GKG  +R I E+A+    +   ++ +D D+ S+K  W+E++   +  G  D
Sbjct: 166 SFVYEGLPGKGSAMRAIMEFALKQ--DAEAVVFLDADLRSVKPWWVERLAGPVLKGEADY 223

Query: 119 IIPNYARKSFEGNATNHFIYPMLVKIFKRDMPYQCISGDFGFSRGLIKDLTLKCN--WHK 176
           + P Y R  F+G  TN+  +PM   ++ + +  Q I GDFG  R L++    K    W+
Sbjct: 224 VTPFYLRHRFDGTITNNVCFPMTAVLYGKKVR-QPIGGDFGVGRKLLEIYLGKPKEIWNT 282

Query: 177 YTLGYGIDIFLTLTAILKSYKIKEIDLQSKIH--KKSFEKIEKIFLEVSQSFFETINDNS 234
               +GIDI++T TAI +S ++ +  L +K+H  K   + ++ +FL+V  + FE +
Sbjct: 283 DVARFGIDIWMTTTAINESGRVVQAALGTKVHDVKDPGKHLKGMFLQVVGTLFELV---- 338

Query: 235 LNQDKLRLNINFESHSRQFIKSSDI------------LSSNDIENLKLRALFLLQEEKQY 282
                    I +E+  ++  K  D+              S DI NLK  A   L+E +
Sbjct: 339 ---------ITYENVWKEIWKIEDVPIYGETPQEEVPSMSIDIGNLKKLARETLEEVEYI 389

Query: 283 LHG-LSEVEWDGI--LSNTINNIYRYSSEEHSL-------YLLPLYLLRVYNY 325
             G LSEV+  G   LS+ ++ +YR + +            LLP Y  R   +
Sbjct: 390 DRGILSEVKESGTLSLSSWVDTLYRSAVQYRKTRDKKVVENLLPFYFARTARF 442
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2519**

A DNA sequence (GASx542) was identified in *S.pyogenes* <SEQ ID 7537> which encodes the amino acid sequence <SEQ ID 7538>. Analysis of this protein sequence reveals the following:

```
Possible site: 20

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -5.31    Transmembrane    3 -  19 (   1 -  21)

----- Final Results -----
               bacterial membrane --- Certainty=0.3123(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

```
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB07427 GB:AP001519 nucleotide sugar dehydrogenase [Bacillus halodurans]
 Identities = 184/388 (47%), Positives = 274/388 (70%), Gaps = 3/388 (0%)

Query: 1   MKITVVGIGYVGLSIGLLLAKEHDVTFFDIDNKKIDLINKRQSPLKEAAINKLLC-KAKN 59
            M IT+ G GYVGLS  +LLA+ +DV  +DI  +K+D+IN R+SP+ +  I + L  K  N
Sbjct: 1   MNITIAGTGYVGLSNAVLLAQHNDVIAYDIVQEKVDMINNRKSPIVDREIEEFLATKELN 60

Query: 60  INATSSEELAYKDATFIILSLPTNL--KFNKLDTSIIEISVSNILKINKKATIVIKSTVP 117
            + AT+ +E A+KDA F+++S PTN   + N  DTS +E  +S++L IN  A +VIKST+P
Sbjct: 61  LTATTDKEKAFKDAQFVVISTPTNYDPEKNYFDTSSVEAVISDVLSINPNAVMVIKSTIP 120

Query: 118 IGFTEYLRNRFHYNDIIFSPEFLREGSTIHDQLYPSRTIVGNESRNSQLFLDILTDISVE 177
            +G+T  +  RF+  +IIFSPEFLREGS ++D L+PSR +VG  ++ +++F  +L   +++
Sbjct: 121 VGYTREVNERFNTKNIIFSPEFLREGSALYDNLHPSRIVVGERTQRAKIFAALLVQGAIK 180

Query: 178 KDSPSLLVGSSEAEAIKLFSNAYLAQKIAFFNELDTFAEMQNLDSKKIIEAMGYDQRIGN 237
            ++   L   S+EAEAIKLF+N YLA ++AFFNELD++AE++ LD+K+II+ +G D RIG
Sbjct: 181 ENIDVLFTDSTEAEAIKLFANTYLAMRVAFFNELDSYAELKGLDAKQIIDGVGLDPRIGT 240

Query: 238 SHNNPSFGFGGYCLPKDIKQLEYHFKEIPAPIITSISESNLLRKIHIAKMILNSSAKTIG 297
             +NNPSFG+GGYCLPKD KQL  +F+++P  II +I ++N  RK H+A MIL    K +G
Sbjct: 241 HYNNPSFGYGGYCLPKDTKQLLANFEDVPNNIIGAIVDANDTRKDHVANMILKREPKVVG 300

Query: 298 IYRINSKKDSDNCRESSTIDVAKLLKSSGKDVIIFEPLINQKKFLGCPLSNDFNEFIKYS 357
            IYR+  K  SDN R+S+ +DV   L ++G +V+++EP ++  +F G  +  DF EF K S
Sbjct: 301 IYRLTMKTGSDNFRQSAILDVMTRLNNAGAEVVVYEPALDATEFDGSKVIEDFAEFKKMS 360

Query: 358 DIIVANRIDDALRKCNSKVFTRDIFQYD 385
            D+IVANR+ D L++   KV+TRD++  D
Sbjct: 361 DVIVANRLSDDLKEVAEKVYTRDLYTRD 388
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2520

A DNA sequence (GASx544R) was identified in *S.pyogenes* <SEQ ID 7539> which encodes the amino acid sequence <SEQ ID 7540>. Analysis of this protein sequence reveals the following:

```
Possible site: 34

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.06    Transmembrane   46 -  62 (  46 -  62)

----- Final Results -----
               bacterial membrane --- Certainty=0.1022(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2521

A DNA sequence (GASx545R) was identified in *S.pyogenes* <SEQ ID 7541> which encodes the amino acid sequence <SEQ ID 7542>. Analysis of this protein sequence reveals the following:

```
Possible site: 58

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.49   Transmembrane  186 - 202 ( 186 - 203)

----- Final Results -----
               bacterial membrane --- Certainty=0.1595(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2522

A DNA sequence (GASx546R) was identified in *S.pyogenes* <SEQ ID 7543> which encodes the amino acid sequence <SEQ ID 7544>. Analysis of this protein sequence reveals the following:

```
Possible site: 47

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2422(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2523

A DNA sequence (GASx547R) was identified in *S.pyogenes* <SEQ ID 7545> which encodes the amino acid sequence <SEQ ID 7546>. Analysis of this protein sequence reveals the following:

```
Possible site: 60

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1612(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2524**

A DNA sequence (GASx548) was identified in *S.pyogenes* <SEQ ID 7547> which encodes the amino acid sequence <SEQ ID 7548>. Analysis of this protein sequence reveals the following:

```
Possible site: 44

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5156(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2525**

A DNA sequence (GASx549R) was identified in *S.pyogenes* <SEQ ID 7549> which encodes the amino acid sequence <SEQ ID 7550>. Analysis of this protein sequence reveals the following:

```
Possible site: 21

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2526**

A DNA sequence (GASx552) was identified in *S.pyogenes* <SEQ ID 7551> which encodes the amino acid sequence <SEQ ID 7552>. Analysis of this protein sequence reveals the following:

```
Possible site: 15

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.59   Transmembrane   83 -  99 (  83 -  99)

----- Final Results -----
               bacterial membrane --- Certainty=0.1235(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2527**

A DNA sequence (GASx553) was identified in *S.pyogenes* <SEQ ID 7553> which encodes the amino acid sequence <SEQ ID 7554>. Analysis of this protein sequence reveals the following:

```
Possible site: 49

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2781(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2528**

A DNA sequence (GASx554) was identified in *S.pyogenes* <SEQ ID 7555> which encodes the amino acid sequence <SEQ ID 7556>. Analysis of this protein sequence reveals the following:

```
Possible site: 18

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2792(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2529**

A DNA sequence (GASx555) was identified in *S.pyogenes* <SEQ ID 7557> which encodes the amino acid sequence <SEQ ID 7558>. Analysis of this protein sequence reveals the following:

```
Possible site: 35

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.00    Transmembrane   49 -  65 (  49 -  65)

----- Final Results -----
               bacterial membrane --- Certainty=0.1001(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAA36631 GB:AB016282 ORF25 [bacteriophage phi-105]
 Identities = 43/118 (36%), Positives = 69/118 (58%), Gaps = 2/118 (1%)

Query: 3    LLDLIGRKRARDKPQNSYEGQDFSYLFG--RTTSGENVDEFKTMQTTAVYACVRVLAEAV 60
             LL+ +  KR+               +FG  +T SGE V E  ++    ++ACV VL++ +
Sbjct: 2    LLERMFEKRSGSSDHEDGFNNILLNMFGGRKTASGERVSESNSLVQPDIFACVNVLSDDI 61

Query: 61   ASLPIHIYERTENGKEKKLDHPLYFLLHDEPNPEMSSFIFRETIMSHLLIWGNAYVQI 118
             A LPIH Y+RT+ G E+K +H     ++  PNP M++F +++ +M+H+L WGNAY  I
Sbjct: 62   AKLPIHTYKRTDGGIERKPEHKSAHAVYARPNPYMTAFTWKKLMMTHVLTWGNAYSYI 119
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2530**

A DNA sequence (GASx556) was identified in *S.pyogenes* <SEQ ID 7559> which encodes the amino acid sequence <SEQ ID 7560>. Analysis of this protein sequence reveals the following:

```
Possible site: 43

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2055(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2531**

A DNA sequence (GASx557) was identified in *S.pyogenes* <SEQ ID 7561> which encodes the amino acid sequence <SEQ ID 7562>. Analysis of this protein sequence reveals the following:

```
Possible site: 50

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1696(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2532**

A DNA sequence (GASx559) was identified in *S.pyogenes* <SEQ ID 7563> which encodes the amino acid sequence <SEQ ID 7564>. Analysis of this protein sequence reveals the following:

```
Possible site: 51

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1556(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB15798 GB:Z99123 alternate gene name: ipa-83d [Bacillus subtilis]
 Identities = 70/263 (26%), Positives = 121/263 (45%), Gaps = 25/263 (9%)

Query: 68  KTIEQIKELK--YSIDAVACWDEALTHIADDISKELGLNPISSLDSQSFRFKDRMRMVCE 125
            + +EQI ++   +  DA+   +E          + LGL       +++ R K++MR
Sbjct: 87  EVVEQIVKVAEMFGADAITTNNELFIAPMAKACERLGLRGAGVQAAENARDKNKMRDAFN 146

Query: 126 AGGLKMPKYKIINQFSDTNKIINW-KYPLIVKPTSFLASIGVKKVYNFSELQQAVSQMLN 184
              G+K  K K +    D    +     PLI+KPT   +SIGV  + +    +   +++ +
Sbjct: 147 KAGVKSIKNKRVTTLEDFRAALEEIGTPLILKPTYLASSIGVTLITDTETAEDEFNRVND 206

Query: 185 VKFPVYIASGVYELGELYNLEPRVLVEEFIDGE-----------EY-SLESVVRNGIYTP 232
               + +   V         E   + EEF+ GE           +Y S+E ++ +G Y P
Sbjct: 207 YLKSINVPKAV-------TFEAPFIAEEFLQGEYGDWYQTEGYSDYISIEGIMADGEYFP 259

Query: 233 LGITKKIVDEKLFMDEIGHIFPSNLNKEEKSRVYSWAEKLHQILQLNHITTHTEFRIGRN 292
            + I  K    ++   E  HI PS L++E K ++   A+K ++ L L +  THTE ++ +N
Sbjct: 260 IAIHDKT--PQIGFTETSHITPSILDEEAKKKIVEAAKKANEGLGLQNCATHTEIKLMKN 317

Query: 293 GDIILIEIGARIGG-DCIPNLMK 314
             +  LIE  AR  G + IPN+ K
Sbjct: 318 REPGLIESAARFAGWNMIPNIKK 340
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2533**

A DNA sequence (GASx561) was identified in *S.pyogenes* <SEQ ID 7565> which encodes the amino acid sequence <SEQ ID 7566>. Analysis of this protein sequence reveals the following:

```
Possible site: 55

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2602(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2534**

A DNA sequence (GASx562) was identified in *S.pyogenes* <SEQ ID 7567> which encodes the amino acid sequence <SEQ ID 7568>. Analysis of this protein sequence reveals the following:

```
Possible site: 34

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAD06696 GB:AE001539 HISTIDYL-TRNA SYNTHETASE [Helicobacter
           pylori J99]
 Identities = 75/309 (24%), Positives = 129/309 (41%), Gaps = 35/309 (11%)

Query: 11  KGYRRQFNQILLGAWGIESAYVDAEIIVATWRGLQRFKGIKVE--FIQLSNKNIFDVLEK 68
           KG  R+F Q      G ES   DAEII      L   K + +E   + ++++ I + + +
Sbjct: 115 KGRYREFTQCDFDFIGSESLVCDAEIIQVIIASL---KALDLEDFCVSINHRKILNGICE 171

Query: 69  DLSKKLRFEDISIEAILGKYLCNNDIEIIKCLYEKDKINMELLISLISKISNKLVKQEFI 128
                   E + I   L K   N   E +K   + D   ++ L+ ++    N L   EF
Sbjct: 172 YFGIAQVNEVLRIVDKLEKIGLNGVEEELKKECDLDSNTIKDLLEMVQIKQNDLSHAEFF 231

Query: 129 -KVLVLYEYVKNFLP----VDCIYFSLS------NLY--------GTGHYSSMNYKIFIR 169
            K+  L +Y +N        ++ +Y  L       NLY        G G+Y+ + Y+  +
Sbjct: 232 EKIAYLKDYNENLKKGIQDLERLYQLLGDLQISQNLYKIDFSIARGLGYYTGIVYETTLN 291

Query: 170 TKSGDIFDIADGGRIDDMVSKFNKVNVLGVCMGIGTTVLSQEI-------EYEIEDRIMI 222
                +  +  GGR D +   F+K N+ GV   IG   L   +       E   + +++I
Sbjct: 292 DMKS-LGSVCSGGRYDHLTKNFSKENLQGVGASIGIDRLIVALSEMQLLDERSTQAKVLI 350

Query: 223 LVEKIDVKIYKNCLELANKLSGYHCSVFEFPYKKIKKFFKHELYSRHHYIIVRLDGSMEY 282
                +   Y N L  + + SG    V+     +KIKK F +  + H ++ V   G  E+
Sbjct: 351 ACMHEEYFSYANRLAESLRQSGIFSEVYP-EAQKIKKPFSYANHKGHEFVAV--IGEEEF 407

Query: 283 RFSSVALKN 291
           +  +++LKN
Sbjct: 408 KSETLSLKN 416
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2535**

A DNA sequence (GASx564) was identified in *S.pyogenes* <SEQ ID 7569> which encodes the amino acid sequence <SEQ ID 7570>. Analysis of this protein sequence reveals the following:

```
Possible site: 56

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1264(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2536

A DNA sequence (GASx576) was identified in *S.pyogenes* <SEQ ID 7571> which encodes the amino acid sequence <SEQ ID 7572>. Analysis of this protein sequence reveals the following:

```
Possible site: 60

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2537

A DNA sequence (GASx577R) was identified in *S.pyogenes* <SEQ ID 7573> which encodes the amino acid sequence <SEQ ID 7574>. Analysis of this protein sequence reveals the following:

```
Possible site: 17

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.60   Transmembrane    2 -  18 (   1 -  18)

----- Final Results -----
               bacterial membrane --- Certainty=0.2041(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2538

A DNA sequence (GASx579) was identified in *S.pyogenes* <SEQ ID 7575> which encodes the amino acid sequence <SEQ ID 7576>. Analysis of this protein sequence reveals the following:

```
Possible site: 13

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3161(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB12286 GB:Z99106 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 62/140 (44%), Positives = 88/140 (62%), Gaps = 3/140 (2%)

Query: 3    LTNYVQEVSLADFGKPLHHKAYWNKRLKTTGGRFFPKDGHLDFNPRMLEEHGELIFRKIV 62
            L    +++S   F KP  H+A +N RLKTTGGR+     +++ N + L EHG      I+
Sbjct: 6    LQKLTEDISETYFKKPFRHQALFNDRLKTTGGRYLLTSHNIELNRKYLIEHGREELIGII 65

Query: 63   RHELCHYHLYFEGRGYHHKDRDFKDLLAQVNGLRY---VPTSSKSKTNHHYSCQTCGQVY 119
            +HELCHYHL+ EG+GY H+DRDF+ LL QVN  R+   +   +++K  + Y C TCGQ Y
Sbjct: 66   KHELCHYHLHLEGKGYKHRDRDFRMLLQQVNAPRFCTPLKKKAENKKTYMYICTTCGQQY 125

Query: 120 QRKRRINLAKYVCGNCHGKL 139
             +KR +N  +Y CG C GK+
Sbjct: 126 IKKRAMNPDRYRCGKCRGKI 145
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2539**

A DNA sequence (GASx587R) was identified in *S.pyogenes* <SEQ ID 7577> which encodes the amino acid sequence <SEQ ID 7578>. Analysis of this protein sequence reveals the following:

```
Possible site: 53

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-10.40   Transmembrane   46 -  62 (  39 -  89)
   INTEGRAL    Likelihood = -5.36   Transmembrane   65 -  81 (  63 -  89)

----- Final Results -----
            bacterial membrane --- Certainty=0.5161(Affirmative) < succ>
             bacterial outside --- Certainty=0.0000(Not Clear) < succ>
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2540**

A DNA sequence (GASx590R) was identified in *S.pyogenes* <SEQ ID 7579> which encodes the amino acid sequence <SEQ ID 7580>. Analysis of this protein sequence reveals the following:

```
Possible site: 35

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
           bacterial cytoplasm --- Certainty=0.2036(Affirmative) < succ>
            bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2541

A DNA sequence (GASx592R) was identified in *S.pyogenes* <SEQ ID 7581> which encodes the amino acid sequence <SEQ ID 7582>. Analysis of this protein sequence reveals the following:

```
Possible site: 23

>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -4.62    Transmembrane    25 -  41 (  24 -  43)

----- Final Results -----
              bacterial membrane --- Certainty=0.2848(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2542

A DNA sequence (GASx600) was identified in *S.pyogenes* <SEQ ID 7583> which encodes the amino acid sequence <SEQ ID 7584>. Analysis of this protein sequence reveals the following:

```
Possible site: 24

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -2.18    Transmembrane     3 -  19 (   2 -  19)

----- Final Results -----
              bacterial membrane --- Certainty=0.1871(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2543

A DNA sequence (GASx603R) was identified in *S.pyogenes* <SEQ ID 7585> which encodes the amino acid sequence <SEQ ID 7586>. Analysis of this protein sequence reveals the following:

```
Possible site: 48

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.3027(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```

```
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA03927 GB:AJ000109 gluthatione peroxidase [Lactococcus lactis]
 Identities = 79/133 (59%), Positives = 103/133 (77%)

Query: 1   VVLVVNTATKCGLTPQYQALQALYDTYHDKGFEVLDFPCNQFLNQAPGDAEEINHFCSLT 60
           VV+VVNTA+KCG TPQ++ L+ LY+TY D+G E+L FPCNQF NQ  G+  EIN FC L
Sbjct: 25  VVIVVNTASKCGFTPQFEGLEKLYETYKDQGLEILGFPCNQFANQDAGENTEINEFCQLN 84

Query: 61  YHTTFPRFAKIKVNGKDADPLFTWLKEEKSGPLGKRIEWNFTKFLIDQNGQVIKRYSSKT 120
           Y  TF  F KIKVNGK+A PL+ +LK+E  G L   I+WNFTKFLID++GQVI+R++ KT
Sbjct: 85  YGVTFTMFQKIKVNGKEAHPLYQFLKKEAKGALSGTIKWNFTKFLIDRDGQVIERFAPKT 144

Query: 121 DPKLIEEDLKALL 133
           +P+ +EE++K LL
Sbjct: 145 EPEEMEEEIKKLL 157
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2544**

A DNA sequence (GASx605) was identified in *S.pyogenes* <SEQ ID 7587> which encodes the amino acid sequence <SEQ ID 7588>. Analysis of this protein sequence reveals the following:

```
Possible site: 26

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3687(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2545**

A DNA sequence (GASx608R) was identified in *S.pyogenes* <SEQ ID 7589> which encodes the amino acid sequence <SEQ ID 7590>. Analysis of this protein sequence reveals the following:

```
Possible site: 17

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1327(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2546

A DNA sequence (GASx616) was identified in *S.pyogenes* <SEQ ID 7591> which encodes the amino acid sequence <SEQ ID 7592>. Analysis of this protein sequence reveals the following:

```
Possible site: 21

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2547

A DNA sequence (GASx617R) was identified in *S.pyogenes* <SEQ ID 7593> which encodes the amino acid sequence <SEQ ID 7594>. Analysis of this protein sequence reveals the following:

```
Possible site: 36

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0677(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2548

A DNA sequence (GASx622R) was identified in *S.pyogenes* <SEQ ID 7595> which encodes the amino acid sequence <SEQ ID 7596>. Analysis of this protein sequence reveals the following:

```
Possible site: 16

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -7.32   Transmembrane    4 -  20 (   1 -  26)

----- Final Results -----
               bacterial membrane --- Certainty=0.3930(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2549

A DNA sequence (GASx632) was identified in *S.pyogenes* <SEQ ID 7597> which encodes the amino acid sequence <SEQ ID 7598>. Analysis of this protein sequence reveals the following:

```
Possible site: 31

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -3.40   Transmembrane   83 -  99 (  82 - 102)
   INTEGRAL    Likelihood = -1.28   Transmembrane  108 - 124 ( 108 - 124)

----- Final Results -----
               bacterial membrane --- Certainty=0.2359(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2550

A DNA sequence (GASx638) was identified in *S.pyogenes* <SEQ ID 7599> which encodes the amino acid sequence <SEQ ID 7600>. Analysis of this protein sequence reveals the following:

```
Possible site: 25

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -0.64   Transmembrane   12 -  28 (  12 -  28)

----- Final Results -----
               bacterial membrane --- Certainty=0.1256(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2551

A DNA sequence (GASx652R) was identified in *S.pyogenes* <SEQ ID 7601> which encodes the amino acid sequence <SEQ ID 7602>. Analysis of this protein sequence reveals the following:

```
Possible site: 16

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2622(Affirmative) < succ>
```

```
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA74610 GB:Y14232 hypothetical protein [Bacteriophage TP901-1]
 Identities = 225/485 (46%), Positives = 308/485 (63%), Gaps = 20/485 (4%)

Query: 2   RKVAIYSRVSTINQAEEGYSIQGQIEALTKYCEAMEWKIYKNYSDAGFSGGKLERPAITE 61
            +KVAIY+RVST NQAEEG+SI  QI+ LTKY EAM W++   Y+DAGFSG KLERPA+
Sbjct: 3   KKVAIYTRVSTTNQAEEGFSIDEQIDRLTKYAEAMGWQVSDTYTDAGFSGAKLERPAMQR 62

Query: 62  LIEDGKNNKFDTILVYKLDRLSRNVKDTLYLVKDVFTANNIHFVSLKENIDTSSAMGNLF 121
            LI D +N  FDT+LVYKLDRLSR+V+DTLYLVKDVFT N I F+SL E+IDTSSAMG+LF
Sbjct: 63  LINDIENKAFDTVLVYKLDRLSRSVRDTLYLVKDVFTKNKIDFISLNESIDTSSAMGSLF 122

Query: 122 LTLLSAIAEFEREQIKERMQFGVMNRAKSGKTTAWKTPPYGYRYNKDEKTLSVNELEAAN 181
            LT+LSAI EFERE IKERM  G + RAKSGK+  W    +GY +N+    L +  L+A
Sbjct: 123 LTILSAINEFERENIKERMTMGKLGRAKSGKSMMWTKTAFGYYHNRKTGILEIVPLQATI 182

Query: 182 VRQMFDMIISGCSIMSITNYARDN-FVGN--TWTHVKVKRILENETYKGLVKYREQTFSG 238
            V Q+F   +SG S+  + +   ++  +G    W++  +++ L+N  Y G +K+++  F G
Sbjct: 183 VEQIFTDYLSGISLTKLRDKLNESGHIGKDIPWSYRTLRQTLDNPVYCGYIKFKDSLFEG 242

Query: 239 DHQAIIDEKTYNKAQIALAHRT----DTKTNTRPFQGKYMLSHIAKCGYCGAPLKVCTGR 294
             H+ II  +TY K Q  L  R     +   N RPFQ KYMLS +A+CGYCGAPLK+  G
Sbjct: 243 MHKPIIPYETYLKVQKELEERQQQTYERNNNPRPFQAKYMLSGMARCGYCGAPLKIVLGH 302

Query: 295 AKNDGTRRQTYVCVNKTESLARRSVNNYNNQKICNTGRYEKKHIEKYVIDVLYKLQHDKE 354
             + DG+R   Y C N+      + +  YN+ K C++G Y+  ++E  VID L   Q + +
Sbjct: 303 KRKDGSRTMKYHCANRFPR-KTKGITVYNDNKKCDSGTYDLSNLENTVIDNLIGFQENND 361

Query: 355 YLKKIKKDDN--IIDITPLKKEIEIIDKKINRLNDLYINDLIDLPKLKKDIEELNHLKDD 412
             L KI   +N  I+D +  KK+I  IDKKI + +DLY+ND I + +LK   + L   K
Sbjct: 362 SLLKIINGNNQPILDTSSFKKQISQIDKKIQKNSDLYLNDFITMDELKDRTDSLQAEK-- 419

Query: 413 YNKAIKLNYLDKKNEDSLGML------MDNLDIRKSSYDVQSRIVKQLIDRVEVTMDNID 466
              K +K    + K  DS  +       + ++ I + SYD + +IV  L+ +V+VT DN+D
Sbjct: 420 --KLLKAKISENKFNDSTDVFELVKTQLGSIPINELSYDNKKKIVNNLVSKVDVTADNVD 477

Query: 467 IIFKF 471
            IIFKF
Sbjct: 478 IIFKF 482
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2552**

A DNA sequence (GASx653R) was identified in *S.pyogenes* <SEQ ID 7603> which encodes the amino acid sequence <SEQ ID 7604>. Analysis of this protein sequence reveals the following:

```
Possible site: 48

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.22    Transmembrane   86 - 102 (  86 - 102)

----- Final Results -----
               bacterial membrane --- Certainty=0.1489(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF12707 GB:AF066865 unknown [bacteriophage TPW22]
 Identities = 45/67 (67%), Positives = 53/67 (78%), Gaps = 2/67 (2%)

Query: 57  EKEAVRCPKCKSTNVGFMQQGKKTFSVKKAVAGTLLIG--GTVMGFLGEKGKKQWHCNEC 114
            +K A++CPKCKST+V FMQQGKK FSV KAV G +L G  GT+ GF G+KGKKQWHCN C
Sbjct: 138 DKHAIKCPKCKSTDVVFMQQGKKGFSVGKAVGGAVLTGGIGTLAGFAGKKGKKQWHCNNC 197

Query: 115 SCIFETK 121
              +FETK
Sbjct: 198 GRVFETK 204
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2553**

A DNA sequence (GASx655) was identified in *S.pyogenes* <SEQ ID 7605> which encodes the amino acid sequence <SEQ ID 7606>. Analysis of this protein sequence reveals the following:

```
Possible site: 50

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3956(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB63661 GB:AJ251789 Cro protein [Lactobacillus casei
          bacteriophage A2]
 Identities = 43/76 (56%), Positives = 55/76 (71%)

Query: 26   MTINLKRLKAERIASGMTQCEVAQSMGWKTRTPYAKRENGIVSIGADELAKITLIFGLPI 85
            MT+NLKRL+AERIA GM Q E+A++MGW TR+ YAKRENGI +I A EL K+  I G
Sbjct: 1    MTLNLKRLRAERIAKGMNQDEMAKAMGWHTRSSYAKRENGITTISATELVKMASILGYGT 60

Query: 86   EKIAIFFDKDVPVMER 101
             ++ +FF  +VP  ER
Sbjct: 61   NQLDLFFTNNVPDRER 76
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2554**

A DNA sequence (GASx656) was identified in *S.pyogenes* <SEQ ID 7607> which encodes the amino acid sequence <SEQ ID 7608>. Analysis of this protein sequence reveals the following:

```
Possible site: 34

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4505(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2555**

A DNA sequence (GASx657) was identified in *S.pyogenes* <SEQ ID 7609> which encodes the amino acid sequence <SEQ ID 7610>. Analysis of this protein sequence reveals the following:

```
Possible site: 35

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.6593(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2556**

A DNA sequence (GASx658) was identified in *S.pyogenes* <SEQ ID 7611> which encodes the amino acid sequence <SEQ ID 7612>. Analysis of this protein sequence reveals the following:

```
Possible site: 32

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5244(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2557**

A DNA sequence (GASx660) was identified in *S.pyogenes* <SEQ ID 7613> which encodes the amino acid sequence <SEQ ID 7614>. Analysis of this protein sequence reveals the following:

```
Possible site: 58

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1133(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```

```
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB99331 GB:U67572 purine NTPase [Methanococcus jannaschii]
 Identities = 71/346 (20%), Positives = 154/346 (43%), Gaps = 52/346 (15%)

Query: 8   MSITINKLEIENVK-----RIKAVKIEPSATGLTIIGGNNNQGKTSVLDAIAWAL--GGN 60
            MS+ + ++ + N K     RIK  K      G+  I G N  GK+S+ +A+ +AL   G+
Sbjct: 1   MSMILKEIRMNNFKSHVNSRIKFEK------GIVAIIGENGSGKSSIFEAVFFALFGAGS 54

Query: 61  KYKPSQAMREGSQ---VPPTLKITMSNGLIVERKGKNASLKVIDPNGQ----------KG 107
             +     + +G +   V    ++  +N  I+           +  NG+          K
Sbjct: 55  NFNYDTIITKGKKSVYVELDFEVNGNNYKIIREYDSGRGGAKLYKNGKPYATTISAVNKA 114

Query: 108 GQQLL----DSFVEELAI---NLPKFMDSTPKEKADVLLEIIGVGDQLAELELKEKEIYN 160
              ++L    + F+  + I    + KF+   P EK + + +++G+ D+  +   K  EI
Sbjct: 115 VNEILGVDRNMFLNSIYIKQGEIAKFLSLKPSEKLETVAKLLGI-DEFEKCYQKMGEIVK 173

Query: 161 QRHAIGVIADQKEKFAKEMTYYPDAPKQLVS-ISELIQQHQAILAKNGE-NAQKR--QNV 216
            +         + E+   E+ Y  +  K+L + +S+L ++++ ++  N + N  K+  +++
Sbjct: 174 E------YEKRLERIEGELNYKENYEKELKNKMSQLEEKNKKLMEINDKLNKIKKEFEDI 227

Query: 217 ERIRYDYNQSILEVDRLRKLLADAEAKTNKLSEDLKIANTD------AMDLHDESTAEIE 270
            E++  ++    L  ++    L + +      +++LKI   D      A +  +    E E
Sbjct: 228 EKLFNEWENKKLLYEKFINKLEERKRALELKNQELKILEYDLNTVVEARETLNRHKDEYE 287

Query: 271 ANIADIDEVNRKVRANFDKDKAE-EDAKQQREQYNILTNDIESIRQ 315
               + +DE+ RK+ +   + K+  ED  +  +Q  I+  DIE +++
Sbjct: 288 KYKSLVDEI-RKIESRLRELKSHYEDYLKLTKQLEIIKGDIEKLKE 332
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2558**

A DNA sequence (GASx661) was identified in *S.pyogenes* <SEQ ID 7615> which encodes the amino acid sequence <SEQ ID 7616>. Analysis of this protein sequence reveals the following:

```
Possible site: 28

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1559(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2559**

A DNA sequence (GASx662) was identified in *S.pyogenes* <SEQ ID 7617> which encodes the amino acid sequence <SEQ ID 7618>. Analysis of this protein sequence reveals the following:

```
Possible site: 52
```

```
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3292(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2560**

A DNA sequence (GASx663) was identified in *S.pyogenes* <SEQ ID 7619> which encodes the amino acid sequence <SEQ ID 7620>. Analysis of this protein sequence reveals the following:

```
Possible site: 15

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4867(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2561**

A DNA sequence (GASx664) was identified in *S.pyogenes* <SEQ ID 7621> which encodes the amino acid sequence <SEQ ID 7622>. Analysis of this protein sequence reveals the following:

```
Possible site: 46

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2141(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2562**

A DNA sequence (GASx667) was identified in *S.pyogenes* <SEQ ID 7623> which encodes the amino acid sequence <SEQ ID 7624>. Analysis of this protein sequence reveals the following:

```
Possible site: 59

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2614(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF80834 GB:AF165214 Orf78 [Pseudomonas phage D3]
 Identities = 68/200 (34%), Positives = 109/200 (54%), Gaps = 10/200 (5%)

Query: 12  GLRFGSLTVINRNRNNSKGGNARWNCLCDCGNKTVVI-GSKLRSGYTKSCGCARKNDNAK 70
            GLR G + V     ++  G  + W C CDCGN+ ++  G+ +R+  T SCGC+R +
Sbjct: 8   GLRVGKVVV--EAFSHCAGKASHWVCRCDCGNRVIMRRGNLMRNRTTTSCGCSRFSH--- 62

Query: 71  GYSSTRLYRIWKGMMNRCYNHKNDNYKYYGGKGISICDEWLTFINFRTWSLSNGYKESLT 130
            G + T  Y  W  M++RC N  N  Y  Y G+GI++C+ W+TF NF       G +   T
Sbjct: 63  GMTGTPTYSSWSNMIDRCTNPSNKRYVDYQGRGITVCERWMTFANFLA---DMGERPDAT 119

Query: 131 -IDRINPKGNYTPLNCRWVSMKMQQNNKTNNRYLSYLGQEYTIAEFSEKLNVTYWTVINQ 189
             +DRI+    Y   NCRW +   Q NN   N ++ YLG+  T+++++ +L +   T+ ++
Sbjct: 120 SLDRIDNDAGYFKENCRWATALEQMNNTRRNTFVEYLGRRQTVSQWAGQLGIPECTLRSR 179

Query: 190 LKLGWSVERIVEEARMKNDR 209
            L  GWS+E  +++   K  R
Sbjct: 180 LNRGWSIEDAMQKPISKQRR 199
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2563**

A DNA sequence (GASx668) was identified in *S.pyogenes* <SEQ ID 7625> which encodes the amino acid sequence <SEQ ID 7626>. Analysis of this protein sequence reveals the following:

```
Possible site: 41

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1476(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB75598 GB:AJ271879 putative DNA helicase [uncultured
           eubacterium]
 Identities = 42/168 (25%), Positives = 75/168 (44%), Gaps = 7/168 (4%)

Query: 374 IAGPSKAGKSFALIELSIALAEGQKWLG-WQCEQGKVLYVNLELDRPSALHRFKDVYDAM 432
           +  P  AGKS   ++L+  +A G   LG  +   G V+Y+  E D P+A+H       A
Sbjct: 35  LVSPGGAGKSMLALQLAAQIAGGPDLLGVGELPTGPVIYLPAE-DPPTAIHHRLHALGAH 93
```

```
Query: 433 GLPPANVANIDIWNLRGKTVPMDKLAPKLIRRSLKKNYQA---VIIDPIYKVLTGDENSA 489
                  A  D   ++     +  +       +LK+  +    +I+D + +    +EN++
Sbjct: 94  LSAEERQAVADGLLIQPLIGSLPNIMASNWFEALKRAAEGRRLMILDTLRRFHIEEENAS 153

Query: 490 DQMAHFTNQFDKVATELGCSVIYCHHHSKGS--QGGKKSMDRASGSGV 535
             MA    + + +A + GCS+++ HH SKG+   G       + GS V
Sbjct: 154 GPMAQVIGRMEAIAADTGCSIVFLHHASKGATMMGAGDQQQASRGSSV 201
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2564**

A DNA sequence (GASx669) was identified in *S.pyogenes* <SEQ ID 7627> which encodes the amino acid sequence <SEQ ID 7628>. Analysis of this protein sequence reveals the following:

```
Possible site: 56

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2555(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2565**

A DNA sequence (GASx670) was identified in *S.pyogenes* <SEQ ID 7629> which encodes the amino acid sequence <SEQ ID 7630>. Analysis of this protein sequence reveals the following:

```
Possible site: 54

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2921(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF74082 GB:AF212845 ORF129 [Lactococcus lactis bacteriophage
          ul36]
 Identities = 36/108 (33%), Positives = 63/108 (58%), Gaps = 1/108 (0%)

Query: 8   IEFFLPMDKIPTTTHQQKKVTVINGKPHFYEPESLKNARDKFTSLLAQHVPPSKLDGPIR 67
           ++F   +DK+PTT  QQK +  + GK  FY+    KN   K   +  +     + + P++
Sbjct: 1   MKFEFELDKMPTT-QQQKGIKKVKGKLQFYDRRGTKNYSLKAQLMKNKPKECFEKNVPLK 59

Query: 68  LTVKWLFPKIKGSTNGQYKTTKPDTDNLQKLLKDCMTELGFWNDDAQV 115
           L+V + +   +     Q+KT++PD DNL K L+D MT+L +++DD+Q+
Sbjct: 60  LSVTFFYAIKQKKRWWQWKTSRPDLDNLMKNLQDYMTKLRYYSDDSQI 107
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2566**

A DNA sequence (GASx671) was identified in *S.pyogenes* <SEQ ID 7631> which encodes the amino acid sequence <SEQ ID 7632>. Analysis of this protein sequence reveals the following:

```
Possible site: 33

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4294(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2567**

A DNA sequence (GASx672R) was identified in *S.pyogenes* <SEQ ID 7633> which encodes the amino acid sequence <SEQ ID 7634>. Analysis of this protein sequence reveals the following:

```
Possible site: 15

>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -6.37   Transmembrane  106 - 122 ( 104 - 125)

----- Final Results -----
               bacterial membrane --- Certainty=0.3548(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2568**

A DNA sequence (GASx673) was identified in *S.pyogenes* <SEQ ID 7635> which encodes the amino acid sequence <SEQ ID 7636>. Analysis of this protein sequence reveals the following:

```
Possible site: 56

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4781(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB18697 GB:U38906 ORF22 [Bacteriophage r1t]
 Identities = 78/207 (37%), Positives = 123/207 (58%), Gaps = 2/207 (0%)

Query: 28  EIHRILGIDEVYKAPKRLTDILFDKDSREDIFRQFLKYETDVSYDWFMQYFEEEQADRKN 87
            + + +L +DE      R+ +++FDK  RE+ + + L    D+  D+F  YF    A
Sbjct: 7   QFYDMLNVDEHMNFTNRIQELVFDKKGREEFYSKILNIHHDMGVDFFRDYFMAHSAVSA- 65

Query: 88  KKQDFTPKSVSTLLSKIISGNQYYEVA-VGTGGILIQAWQEQRLNDSPFTYRPSKYWYHV 146
            K Q +TP  +  L + ++ G+   ++   GTG ++IQ WQ+ R+N   F Y PS YWY
Sbjct: 66  KGQHYTPDELGKLTALLVGGSGGADLTGAGTGTLIIQKWQDDRMNTDFFNYLPSNYWYQA 125

Query: 147 EELSDKAVPFLLFNMSIRGINGVVVHGDSLTRQVKNIYFLQNTKDDMLSFSDINVMPRTQ 206
             ELSD+A+ FL+   +IRG+NGVV+HGD+L   VK +YF+QN+ ++ + FS+INV+P ++
Sbjct: 126 LELSDEAISFLIHAFAIRGMNGVVIHGDALEMAVKQVYFIQNSANNPIGFSEINVIPHSK 185

Query: 207 DIEREFNVKEWIGDGIEHIENPLIEWI 233
            D      + EW    IEHIE+   +WI
Sbjct: 186 DAMEFLGIHEWTEQAIEHIESKFPDWI 212
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2569**

A DNA sequence (GASx674) was identified in *S.pyogenes* <SEQ ID 7637> which encodes the amino acid sequence <SEQ ID 7638>. Analysis of this protein sequence reveals the following:

```
Possible site: 51

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.00    Transmembrane  122 - 138 ( 122 - 138)

----- Final Results -----
               bacterial membrane --- Certainty=0.1001(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF63071 GB:AF158600 gp137 [Streptococcus thermophilus
           bacteriophage Sfi11]
 Identities = 66/135 (48%), Positives = 89/135 (65%), Gaps = 2/135 (1%)

Query: 5   PEIDIQKTKSNAKRKLREYPRWRRIANDVDTQKVTATYSFEPRQSHGVPSKPVERLALNR 64
            PEID + T    KRKLREYPRWR IA+D   QK+T  ++F PR   G  +KPVE +A+ R
Sbjct: 4   PEIDEKATLKRCKRKLREYPRWREIAHDSAEQKITQEFTFMPRG--GGVNKPVENIAVRR 61

Query: 65  VSAEQELDAIEQAVSMILEPERRRILYDKYLAPYKKADKVIYTELCMSESFYYDTLDIAL 124
            V A  EL+AIEQAV+ +  P+ RRIL +KYLA   K +  I   +    + + + L+ ++
Sbjct: 62  VDALNELEAIEQAVNGLYRPDYRRILIEKYLAYPPKPNWQIAQSIGFERTAFQELLNNSI 121

Query: 125 LAFAELYREGVLLVE 139
            LAFAELYR+G L+VE
Sbjct: 122 LAFAELYRDGRLIVE 136
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2570**

A DNA sequence (GASx675) was identified in *S.pyogenes* <SEQ ID 7639> which encodes the amino acid sequence <SEQ ID 7640>. Analysis of this protein sequence reveals the following:

```
Possible site: 41

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1865(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2571**

A DNA sequence (GASx676) was identified in *S.pyogenes* <SEQ ID 7641> which encodes the amino acid sequence <SEQ ID 7642>. Analysis of this protein sequence reveals the following:

```
Possible site: 31

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4870(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB07254 GB:AP001519 unknown [Bacillus halodurans]
 Identities = 194/451 (43%), Positives = 262/451 (58%), Gaps = 69/451 (15%)

Query: 1   MEFVDKKLSEITPYKNNPRNNDEAVGPVAE----SIKEFGFKVPIVV-DKNGEIVNGHTR 55
           +  V+KK+ ++ P + NPR + +   P  E    SI+EFG   PIV  ++ G +V GH R
Sbjct: 3   IRIVNKKIDDLVPAEYNPRLDLQPGDPEYEKLKRSIEEFGLVEPIVFNERTGRVVGGHQR 62

Query: 56  YKAAQKLGLETVPVIVADDLSEEQIKAFRLADNKV-GEIAVWDLDLLNEELNDILDLDMS 114
            K  ++LG E VPV V D L +   KA  +A NK+ G+   + L  L EEL+  L +D++
Sbjct: 63  LKILRELGWEEVPVSVVD-LDDHHEKALNVALNKIEGDWDNFKLKELLEELDSGL-IDVT 120

Query: 115 AFGFDVLDNLDDL-----IEDEKDL--DDF----TGTVPDEPKSKLGDIYQLGSHKLMCG 163
             GFD  + ++DL     +EDE ++  DDF         +EP +K GD++ LG H L+ G
Sbjct: 121 LTGFDE-EEIEDLMTQFFVEDENEIKEDDFDPDEVAEEIEEPITKPGDLWHLGRHFLLVG 179

Query: 164 DSTNGADVKKLMNGELADLLLTDPPYNVAYEGKTKDSLTIKNDSMDNDSFRQFLVNAFSS 223
           DST   DVK+LM  E AD++ TDPPYNV YEG T   + IKND+M++  F QFL +AF +
Sbjct: 180 DSTKIEDVKRLMGNEKADMIFTDPPYNVDYEGAT--GMKIKNDNMEDSEFYQFLFDAFVA 237

Query: 224 ANEVMKPGAVFYIWHADSEGYNFRGACFDIGWTVRQCLIWNKNSMVLGRQDYHWKHEPCL 283
             +V K G   Y+ HADSEG  FR A  D G+ ++QCLIW KNS+VLGRQDYHW+HEP L
Sbjct: 238 MYQVTKEGGPIYVCHADSEGLTFRKAFQDSGFLLKQCLIWVKNSLVLGRQDYHWRHEPIL 297

Query: 284 YGWKDGAGHLWASDRKQTSVID-------------------------------------- 305
           YGWK GA H W   RKQ++VI+
Sbjct: 298 YGWKPGAAHKWYGGRKQSTVIEDPVDLAITPKVDHVLLTFNNGISSTVVKVPSYEIIHDG 357
```

```
Query: 306 ---------YEKPQRNGVHPTMKPVGLFDYQIKNNTKGSDIVLDLFGGSGTTLIACESNG 356
                     E+P+RN  HPTMKP+ L    I+N++K  + VLD FGGSG+TLIACE  G
Sbjct: 358 SDEGMTTWRIERPKRNADHPTMKPIALCARAIQNSSKPGERVLDPFGGSGSTLIACEQTG 417

Query: 357 RHARLMEYDPKYVDVIIKRWEELTGESVIQL 387
           R   +MEYDP Y +VII+RWEE TG++ ++L
Sbjct: 418 RICHMMEYDPVYAEVIIRRWEEWTGQNAVKL 448
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2572**

A DNA sequence (GASx677) was identified in *S.pyogenes* <SEQ ID 7643> which encodes the amino acid sequence <SEQ ID 7644>. Analysis of this protein sequence reveals the following:

```
Possible site: 54

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4744(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2573**

A DNA sequence (GASx678) was identified in *S.pyogenes* <SEQ ID 7645> which encodes the amino acid sequence <SEQ ID 7646>. Analysis of this protein sequence reveals the following:

```
Possible site: 31

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.27   Transmembrane   90 - 106 (  90 - 106)

----- Final Results -----
               bacterial membrane --- Certainty=0.1107(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2574**

A DNA sequence (GASx679) was identified in *S.pyogenes* <SEQ ID 7647> which encodes the amino acid sequence <SEQ ID 7648>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
```

```
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3408(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA66734 GB:X98106 minor capsid protein [Bacteriophage phig1e]
 Identities = 213/494 (43%), Positives = 323/494 (65%), Gaps = 19/494 (3%)

Query: 1   MGVIQKIKNLVTRSKYVM-TTQSLTNITDHPKIAISKLEYDRITTNLKYYKSDWDSVLYL 59
           MG+IQ+IK+L  +       T SL+ ITD P+I+I   EY RI T+L YY     + Y
Sbjct: 1   MGLIQRIKDLFWKGAAATGVTGSLSKITDDPRISIDPDEYVRIQTDLDYYSDKLQYIHYQ 60

Query: 60  NTDGETKKRDLNHLPIARTAAKKIASLVFNEQAEIKV-DDDAANEFISETLKNDRFNKNF 118
            +DG  KKR  N + +A+TAA++IAS+VFNE+AEI V D++ A++F+++ L+++ F   F
Sbjct: 61  ASDGIKKKRLKNTINMAKTAARRIASVVFNEKAEIHVKDNNEADKFLNDVLEDNDFKNKF 120

Query: 119 ERYLESCLALGGLAMRPYVDGDKVRVAFVQAPVFLPLQSNTQDVSSAAVVIKSVKTINGK 178
           E  LE  +ALGG AMRPY+DG+ +++A+V+A  F PLQSNT D+S AA+  ++ +T + +
Sbjct: 121 EEALEKGVALGGFAMRPYIDGNHIKIAWVRADQFYPLQSNTNDISEAAIASRTQRTESNQ 180

Query: 179 EVYYTLIEFHEWQSSDDYVISNELYRSDDKAKVGSRVPLS--EVYKDLKDEAKVTDVTRP 236
             YYTL+EFH+WQ +  Y I+NELY+SD    VG++VPLS   VYK+L  +  ++ + RP
Sbjct: 181 TKYYTLLEFHQWQDNGSYQITNELYKSDSPDIVGNQVPLSTLPVYKELAPQVTISGLQRP 240

Query: 237 IFTYLKTPGMNNKDINSPLGLSIFDNAKTTIDFINTTYDEFMWEVKMGQRRVAVPESLTA 296
           +F Y KTPG NN +I SPLGL + DNAK  +D IN T+D+F+WE+++GQ+ +AV   +
Sbjct: 241 LFAYFKTPGANNINIESPLGLGVVDNAKHVLDDINDTHDQFIWEIRLGQKHIAVQPGMLR 300

Query: 297 LTVRTADGDVVPRPRFESDQNVYIRMGGRDLDSSAIQDLTTPIRADDYIKAINEGLSLFE 356
                   D   +P F+++QNVY+ +   D +   ++D+TTPIR   Y  AI+  +  FE
Sbjct: 301 F-------DDEHKPTFDTEQNVYVGVLSDDNNGLGVKDMTTPIRTVQYKDAIDHFIKEFE 353

Query: 357 MQIGVSAGLFSFDGKSMKTATEIVSENSDTYQMRNSIVTLVEQSLKELVISIFEIAKAYD 416
           +QIG+S G FS+    +KTATE+VS NS TYQ R+S +T+VE+++ EL  SIFE+A A
Sbjct: 354 VQIGLSTGTFSYSNDGVKTATEVVSNNSMTYQTRSSYLTMVEKAIDELCQSIFELANAGA 413

Query: 417 LYQSEVP--SMDNISISL------DDGVFTDRDAELDYWIKVVNAGFGTREMAIQKVLNV 468
           L+    P  ++D+ S  L      DDGVF ++D +L+   KV+  G  +++  +Q+   +
Sbjct: 414 LFDDGKPLFTLDSASQPLDIECHFDDGVFVNKDKQLEEDAKVLAIGALSKQTFLQRNYGM 473

Query: 469 TEEKAQEIAAEINT 482
           T+E+A E  A+I +
Sbjct: 474 TDEQAAEELAKIQS 487
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2575**

A DNA sequence (GASx680) was identified in *S.pyogenes* <SEQ ID 7649> which encodes the amino acid sequence <SEQ ID 7650>. Analysis of this protein sequence reveals the following:

```
Possible site: 48

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1840(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```

```
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB53790 GB:AJ242593 gp4 [Bacteriophage A118]
 Identities = 114/385 (29%), Positives = 187/385 (47%), Gaps = 23/385 (5%)

Query: 8   LNDEQLLLEASQLSDMYHQLTLDLFDQVIERIKARGSASLADNPYLWQANKLHDVGLLNA 67
           L   QL L    + D+Y  L  +LF  ++ R+K + + S ADN   WQ  KL+ V  L+
Sbjct: 3   LTPRQLDLFVQPIVDVYTGLENELFTLIVRRLKTKKNIS-ADNVLAWQIEKLNQVHALDQ 61

Query: 68  DNIKLIAKYSGIAEAQLRYIIKNEGFKIYKNTSEQLEEALGRESGV-------NSTIQDD 120
             I+ I+K SG++  +L  ++K+ G+   K       +    E+G          TI D
Sbjct: 62  QMIERISKASGVSAKKLFSVVKDAGYSDLKQVDNYFSKLA--EAGAVLPLVSDGQTIVDK 119

Query: 121 LSNYARQAIDDVHNLTNTTLPFSVIGAYQGIIQDAVAGVVTGLKTPDQAINQTVIKWFKK 180
           +     +  +  +   N T+       Y  II +    V+ GLKT  QA+ +TV K+ +
Sbjct: 120 VMRSYFKLAESNYKRINQTMLSQARQIYSDIIHETTQSVLAGLKTHRQALAETVTKFAEN 179

Query: 181 GFYGFTDKAGRKWRADSYARTVINTTTWRVFNEAKEAPAREFGIDTFYYSKKATAREMCA 240
           G     DKA ++W  ++Y RTV  TT   V+N  ++    E+G+D    S+   AR  C+
Sbjct: 180 GVPALVDKANKRWTPEAYVRTVTRTTVNSVYNSVEDERMNEYGVDLVRISQHVGARPTCS 239

Query: 241 PLQHQIV---TTGEAREEGGIKILALSD----YGHGEPDGCLGINCKHTKTPFVVGVNSK 293
            +Q +++   +  E R + G K +++      YG+G  DG  G NC+H +  F+ G+N
Sbjct: 240 IVQGKVICLLSVEETRSKYGNKYMSIYSPELRYGYG--DGIFGCNCRHHRFAFIEGINIA 297

Query: 294 PELPEHLKNITPAQAKANANAQAKQRAIERSIRKSKELLHVAKQLGDKELIRQYQSDVRS 353
           P+  E    I   + K       +QR +ER IR +K  L  A++LGD+  +++ +  VR+
Sbjct: 298 PDESE---LIDEEENKRVYALSQQQRLMERDIRAAKRKLSAAEELGDELAVKKAKQAVRT 354

Query: 354 KQDALNYLINNNAFLHRNQAREKRY 378
           KQ  L   +  +  L R  +REK Y
Sbjct: 355 KQSKLRAFVKTHN-LTRQYSREKVY 378
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2576**

A DNA sequence (GASx681) was identified in *S.pyogenes* <SEQ ID 7651> which encodes the amino acid sequence <SEQ ID 7652>. Analysis of this protein sequence reveals the following:

```
Possible site: 31

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2756(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2577

A DNA sequence (GASx682) was identified in *S.pyogenes* <SEQ ID 7653> which encodes the amino acid sequence <SEQ ID 7654>:

```
TLDNQSVIKAIGDTVDYIKKNYKRKWGK
```

Analysis of this protein sequence reveals the following:

```
Possible site: 25

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2578

A DNA sequence (GASx683) was identified in *S.pyogenes* <SEQ ID 7655> which encodes the amino acid sequence <SEQ ID 7656>. Analysis of this protein sequence reveals the following:

```
Possible site: 60

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.5288(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2579

A DNA sequence (GASx685) was identified in *S.pyogenes* <SEQ ID 7657> which encodes the amino acid sequence <SEQ ID 7658>:

```
GATEVGANRVVSGVYGEVLGVQIVRSRKCPKGTAYMVRKGALRIMLKRNTMVETDRDITKAINQIVANKHYGVYLYKAEKAVKITLKDAAK
K
```

Analysis of this protein sequence reveals the following:

```
Possible site: 18

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1750(Affirmative) < succ>
```

```
            bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA59185 GB:X84706 major head protein [Bacteriophage B1]
 Identities = 138/270 (51%), Positives = 186/270 (68%), Gaps = 6/270 (2%)

Query: 1   MAVGTTKMAQMLDPEVLADMIDAEVGKAIRFAPLAEVDTTLEGQPGTTLTVPK-WDYIGD 59
            M+   T +A +++PEVLA ++  E+ KA+RFAPLA+VDTTL+GQPG TL  P  + YIGD
Sbjct: 1   MSKQKTTLADLVNPEVLATIVSYELNKALRFAPLAQVDTTLQGQPGNTLKFPDPFTYIGD 60

Query: 60  AEDVAEGEAIPMTQLGFKKTTMTIKKAGKGVEITDEAILSGYGDPVGQAAKQIVEAIDHK 119
            A DVAEG  I + ++G    ++TIKKA KG EITDEA LSGYGDP+G++ KQ+  ++ +K
Sbjct: 61  AADVAEGGEISLDKIGTTTKSVTIKKAAKGTEITDEAALSGYGDPIGESNKQLGLSLANK 120

Query: 120 VDADVLDALSKSTQTVEATATVDGVSKALDIFNDEDDAETVIVMNPADASTLRLDAAKEW 179
            VD D+L A   ++QTV   A VDGV  ALDIFNDED    V+++NP DA+ +R DA  +
Sbjct: 121 VDDDLLSAAKTTSQTVSTKANVDGVQAALDIFNDEDAQAYVLIVNPKDAAKIRKDANAKN 180

Query: 180 LGATEVGANRVVSGVYGEVLGVQIVRSRKCPKGTAYMVR----KGALRIMLKRNTMVETD 235
            +G +EVGAN +++G Y +VLG QIVRS+K  +G+A M +      AL+++LKR   VETD
Sbjct: 181 IG-SEVGANALINGTYADVLGAQIVRSKKLAEGSALMFKIVSNSPALKLVLKRGVQVETD 239

Query: 236 RDITKAINQIVANKHYGVYLYKAEKAVKIT 265
            RDI      I A++HY  YLY   K V IT
Sbjct: 240 RDIVTKTTVITADEHYAAYLYDLTKVVNIT 269
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2580**

A DNA sequence (GASx686) was identified in *S.pyogenes* <SEQ ID 7659> which encodes the amino acid sequence <SEQ ID 7660>. Analysis of this protein sequence reveals the following:

```
Possible site: 35

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2581**

A DNA sequence (GASx687) was identified in *S.pyogenes* <SEQ ID 7661> which encodes the amino acid sequence <SEQ ID 7662>. Analysis of this protein sequence reveals the following:

```
Possible site: 54

>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2942(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2582**

A DNA sequence (GASx688) was identified in *S.pyogenes* <SEQ ID 7663> which encodes the amino acid sequence <SEQ ID 7664>. Analysis of this protein sequence reveals the following:

```
Possible site: 21

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2844(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC00538 GB:L02496 unknown protein [Bacteriophage LL-H]
 Identities = 35/86 (40%), Positives = 48/86 (55%), Gaps = 6/86 (6%)

Query: 24   KLIMNNQVMMSMNPYVPYRDGALRGSSRANSVGVTWSGPHARAQFYGGAYNKYKSFKFKK 83
             +L + NQ+   M  YVP R G LR  S  N  G+ ++  +ARAQFYG         + +
Sbjct: 20   RLQVLNQMHQDMEQYVPKRAGFLRSQSFVNDTGIHYTAKYARAQFYGFV----NGHRVRN 75

Query: 84   YTTPGTGKRWDKRALANATIVKDWEK 109
             Y+TPGTG+RWD +  A A    DW+K
Sbjct: 76   YSTPGTGRRWDLK--AKAVYKADWQK 99
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2583**

A DNA sequence (GASx689) was identified in *S.pyogenes* <SEQ ID 7665> which encodes the amino acid sequence <SEQ ID 7666>. Analysis of this protein sequence reveals the following:

```
Possible site: 45

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2892(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA66741 GB:X98106 minor capsid protein [Bacteriophage phig1e]
```

```
 Identities = 36/109 (33%), Positives = 64/109 (58%), Gaps = 2/109 (1%)

Query: 17  DLGIKPRLDYLTRQEDLAIYPMPGGKVNNEYMDGTREISLPFEIAIKTKNQELASTVMWT 76
           +L +K  L YLT  + L++YP+PG +V +E   G ++  + +E+ ++TKNQ+ A+T +W
Sbjct: 16  NLPMKCTLGYLTAADSLSLYPLPGSRVLDEDYAGNQQWQMNYEVGMRTKNQQQANTTLWL 75

Query: 77  INSALSNFDL-KLPSLNHSYTFISLDVE-KPFLNDLSDQGFYIYVLDIT 123
           ++ AL       L S N S+ F SL +  +P +++   QG+  Y L  +
Sbjct: 76  VSQALDVLTADDLVSSNGSFEFESLTINGQPSISEQDTQGYSTYQLSFS 124
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2584**

A DNA sequence (GASx690) was identified in *S.pyogenes* <SEQ ID 7667> which encodes the amino acid sequence <SEQ ID 7668>. Analysis of this protein sequence reveals the following:

```
Possible site: 18

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1626(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB53798 GB:AJ242593 major tail shaft protein [Bacteriophage
           A118]
 Identities = 54/133 (40%), Positives = 77/133 (57%), Gaps = 9/133 (6%)

Query: 1   MRQKNALRGHFIAPYVKGEEKTEVTKEKLLELARWIKDISDDTDEKTEDEAYYDGDGTEE 60
           MR KNA   + +A  V G  + +  +     L++WI ++SDD  + TE++  YDGDG E+
Sbjct: 1   MRIKNAKTKYSVAEIVAGAGEPDWKR-----LSKWITNVSDDGSDNTEEQGDYDGDGNEK 55

Query: 61  TTVVGVKGAYTFEGTYDPEDKAQAHIASLKYKLGDERKVWHLIVSADGKTQWLGVATVTE 120
           T V+G   AYTFEGT+D ED+AQ  I + K +  + R +   I   D +T  +G ATV+E
Sbjct: 56  TVVLGYSEAYTFEGTHDREDEAQNLIVA-KRRTPENRSIMFKIEIPDTETA-IGKATVSE 113

Query: 121 I--IAGSGAAARF 131
           I   AG G A  F
Sbjct: 114 IKGSAGGGDATEF 126
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2585**

A DNA sequence (GASx691) was identified in *S.pyogenes* <SEQ ID 7669> which encodes the amino acid sequence <SEQ ID 7670>. Analysis of this protein sequence reveals the following:

```
Possible site: 17

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3521(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2586**

A DNA sequence (GASx692) was identified in *S.pyogenes* <SEQ ID 7671> which encodes the amino acid sequence <SEQ ID 7672>. Analysis of this protein sequence reveals the following:

```
Possible site: 61

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3438(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB53801 GB:AJ242593 gp15 [Bacteriophage A118]
 Identities = 67/191 (35%), Positives = 110/191 (57%), Gaps = 17/191 (8%)

Query: 11  FEFRGEIYPIDLSFNKVLDVFDVIDDDFLNEAEKCFLCLDILLDRTDLPFTYAVD----- 65
            +E+ G+ Y +DL+F+ VL V D+ +D+ L++  +  L +D+L    D+P+  + +
Sbjct: 12  YEYEGKEYKLDLAFDNVLRVIDLTEDNSLSDVFRANLAIDVLF-ADDMPWPRSNEEDEYA 70

Query: 66  -------LWVYIKTNFIDAERPEKPQLDIKGNPMPVVKEKEDNKKVI---DLSLDAEFIY 115
                   + + I TN+I  E  +    DI GN MP      D+ + I    L+ DA++IY
Sbjct: 71  NIEEKSLVLIDIFTNYIVKENDDGLLYDIDGNKMPSATNNNDDAEEIASYSLTQDADYIY 130

Query: 116 ASFRQAYQINLLKEQNRLSWIEFKALLNALPDDTVMQRIIAIRQWE-DDGEGSKKYRDNM 174
            ASF Q Y I+LL  + ++ W +F+ALL +L DDT ++ II IRQ E   G+G++K R+ +
Sbjct: 131 ASFLQDYNIDLLDSRGKMHWYKFRALLESLRDDTTIKTIIGIRQAELPSGKGTEKERNEL 190

Query: 175 RKLKAKYSLDE 185
             KLK +Y L +
Sbjct: 191 IKLKNRYKLKD 201
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2587**

A DNA sequence (GASx694) was identified in *S.pyogenes* <SEQ ID 7673> which encodes the amino acid sequence <SEQ ID 7674>. Analysis of this protein sequence reveals the following:

```
Possible site: 29

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4143(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAG18639 GB:AY007505 unknown [Streptococcus mitis]
 Identities = 48/157 (30%), Positives = 85/157 (53%), Gaps = 10/157 (6%)

Query: 86  DLELSWEPDYIYKATHITPFSIKEVLRNFGRLKINFLIHPIKYLKTGKQEVPLVNG-GTL 144
            +LE S+ P+ ++ A H    S K    +  +LKI   + P +Y KT   E    NG GT+
Sbjct: 81  ELEFSYHPESVFYA-HFLTASYKPFGNHAWQLKIKLNMQPFRYQKTVNPES--YNGPGTI 137

Query: 145 QNPGNVQAKPILKIKGTGNGILTINDFETGLENVQSELVIDMERHLVYKDVLSAWDNIVR 204
             NPG + ++PI++++G G+  +TI   ET   NV+++  ID  +    +++ +A   +
Sbjct: 138 NNPGTIYSEPIIEVQGDGDVSITIGR-ETMYLNVKTKATIDCRQG--RQNIYNATGAVQN 194

Query: 205 TERHRMPLFDV--GQNKISWTGS-FTITAVPNWGVKV 238
            T R R   F++  G++ I++TG+   +   PNW  K+
Sbjct: 195 TLRKRGGFFEIPTGRSGITFTGNVLRLIIRPNWRYKI 231
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2588**

A DNA sequence (GASx695R) was identified in *S.pyogenes* <SEQ ID 7675> which encodes the amino acid sequence <SEQ ID 7676>. Analysis of this protein sequence reveals the following:

```
Possible site: 15

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.60    Transmembrane   15 - 31 (  15 -  31)

----- Final Results -----
              bacterial membrane --- Certainty=0.2041(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2589**

A DNA sequence (GASx697) was identified in *S.pyogenes* <SEQ ID 7677> which encodes the amino acid sequence <SEQ ID 7678>. Analysis of this protein sequence reveals the following:

```
Possible site: 22

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.3348(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAA86895 GB:U28144 hyaluronidase [Streptococcus pyogenes]
 Identities = 326/337 (96%), Positives = 329/337 (96%)
```

```
Query: 1   MSENIPLRVQFKRMKAAEWARSDVILLESEIGFETDTGFARAGDGHNRFSDLGYISPLDY 60
           MSENIPLRVQFKRMKAAEWARSDVILLESEIGFETDTGFARAGDGHNRFSDLGYISPLDY
Sbjct: 1   MSENIPLRVQFKRMKAAEWARSDVILLESEIGFETDTGFARAGDGHNRFSDLGYISPLDY 60

Query: 61  NLLTNKPNIDGLATKVETAQKLQQKADKETVYTKAESKQELDKKLNLKGGVMTGQLKFKP 120
           NLLTNKPNIDGLATKVETAQKLQQKADKETVYTKAESKQELDKKLNLKGGVMTGQLKFKP
Sbjct: 61  NLLTNKPNIDGLATKVETAQKLQQKADKETVYTKAESKQELDKKLNLKGGVMTGQLKFKP 120

Query: 121 AATVAYSSSTGGAVNIDLSSTRGAGVVVYSDNDTSDGPLMSLRTGKETFNQSALFVDYKG 180
           AATVAYSSSTGGAVNIDLSSTRGAGVVVYSDNDTSDGPLMSLRTGKETFNQSALFVDYKG
Sbjct: 121 AATVAYSSSTGGAVNIDLSSTRGAGVVVYSDNDTSDGPLMSLRTGKETFNQSALFVDYKG 180

Query: 181 TTNAVNIAMRQPTTPNFSSALNITSGNENGSAMQLRGSEKALGTLKITHENPSIGADYDK 240
           TTNAVNIAMR  TTPNFSSALNITSGNENGSAMQLRGSEKALGTLKITHENPSIGADYDK
Sbjct: 181 TTNAVNIAMRHATTPNFSSALNITSGNENGSAMQLRGSEKALGTLKITHENPSIGADYDK 240

Query: 241 NAAALSIDIVKKTNGAGTAAQGIYINSTSGTTGKLLRIRNLSDDKFYVKSDGGFYAKETS 300
           NAA   + + K+ NGAGTAAQGIYINSTSGTTGKLLRIRNLSDDKFYVKSDGGFYAKETS
Sbjct: 241 NAARYPLILSKRQNGAGTAAQGIYINSTSGTTGKLLRIRNLSDDKFYVKSDGGFYAKETS 300

Query: 301 QIDGNLKLKDPTANDHAATKAYVDKAISELKKLILKK 337
           QIDGNLKLKDPTANDHAATKAYVDKAISELKKLILKK
Sbjct: 301 QIDGNLKLKDPTANDHAATKAYVDKAISELKKLILKK 337
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2590**

A DNA sequence (GASx698) was identified in *S.pyogenes* <SEQ ID 7679> which encodes the amino acid sequence <SEQ ID 7680>. Analysis of this protein sequence reveals the following:

```
Possible site: 17

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4208(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>

RGD motif 54-56
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAA98102 GB:M19348 ORF [Streptococcus pyogenes phage H4489A]
 Identities = 250/648 (38%), Positives = 351/648 (53%), Gaps = 75/648 (11%)

Query: 1   MSRDPTLILDESNLVIGKDGRVHYTFTTEDDNPKVRLASKCLGTAHFNQLMIERGDQATS 60
           MSRDPT  ++E +L    DGR + TF  +  +  VRL S CLG     +L +E  +
Sbjct: 1   MSRDPTYTINEHDLSFA-DGRFYVTFKADKSSETVRLNSSCLGNTIIKKLQVEDDNTMHD 59

Query: 61  YVAPVVVEGTGNPTGLFKDLKEISLELTDTANSQLWSKIKLTNRGMLQEYYDGKIKTEIV 120
           +V P V   T    GL + +KE+ L+L D   S LW KIK  N+ ML EY + ++ + I
Sbjct: 60  FVKPKVT--TQQAFGLAQQVKELDLQLKDP-KSDLWGKIKFNNKAMLVEYANKEMSSAIA 116

Query: 121 NSARGVATRISEDTDKKLALINDTIDGIRREYRDADRKLSASYQAGIEGLKATMANDKIG 180
            SA  +  ++    D++ +    T++GI++  +
Sbjct: 117 QSAEQILLQVKSIDDERYSKFEQTLNGIKQTVKSES------------------------ 152

Query: 181 LQAEIKASAQGLSQKYDDELRKLSAKITTTSSGTTEAYESKLAGLRAEFTRSNQGTRTEL 240
               ++++   L+  +D  +  L  K +  S  T ++  S+L            G  + L
Sbjct: 153 ----VESARTQLASMFDSRISGLDGKYSRLSQ-TIDSLSSRLD--------DGVGNYSTL 199

Query: 241 ESQISGLRAVQQSTASQISQEIRDREGAVSRVQQSLESYQRRMQDAEENYSSLTHTVRGL 300
```

```
             ++SG           I   + +    VSR+ Q+ +  Q ++ +A +NYSSL+ TV+GL
Sbjct: 200 SQKVSG-----------IDLRVSNAANDVSRLSQTAQGLQSQITNANQNYSSLSQTVQGL 248

Query: 301 QSDVGSPTGKIQSRLTQLAGQIEQRVTRDGVMSIISGAGDSIKLAIQKAGGINAKMSGNE 360
           Q+ V        SR+ QL+  I  +VT+  V + I+ + D I  AI+     + KM+G+E
Sbjct: 249 QTTVRDNQSNATSRINQLSDLISTKVTKGDVETTIAQSYDKIAFAIRDKLPAS-KMTGSE 307

Query: 361 IISAINLNSYGVTIAGKHIALDGNTTVNGTFTTKIAEAIKIRADQIIAGTIDAARIRVIN 420
           IISAINL+  GV I GK+I LDGN+ ++     K A    + A +I  G ++A+RI
Sbjct: 308 IISAINLDRSGVKITGKNITLDGNSYISNA-VIKDAHIANMDAGKINTGYLNASRIAAEA 366

Query: 421 LNASSIVGLDANFIK--AKIGY---------------AIT---DLLEGKVIKARNGAMLI 460
           +    I    A F K  A  GY               A+T     + G V+ A NGA
Sbjct: 367 ITGDKIKMDYAFFNKLTANEGYFRTLFAKNIFTTSVQAVTTSASKITGGVLSATNGASRW 426

Query: 461 DLNTAKMDFNSDATINFNSKNNALVRKDGTHTAFVHFSNATPKGYTGSALYASIGITSSG 520
           DLN+A +DFN DATINFNSKNNALVRK GT+TAFVHFSNATPKGY GSALYASIGITSSG
Sbjct: 427 DLNSANIDFNRDATINFNSKNNALVRKSGTNTAFVHFSNATPKGYRGSALYASIGITSSG 486

Query: 521 DGVNSASSGRFAGLRSFRYATGYNHTAAVDQTEIYGDNVLVVDDFNITRGFKFRPDKMQK 580
           DG++SASSGRF G+R FRYA G  HTA VDQ EIYGD+++  DDFNI RGFK RP  M K
Sbjct: 487 DGIDSASSGRFCGVRFFRYAEGLQHTAKVDQAEIYGDDIVFSDDFNIDRGFKMRPSLMPK 546

Query: 581 MLDMNDLYAAVVALGRCWGHLANVGWNTAHSNFTSAVNRELNNYITKI 628
           M+D+N +Y A++ALGRCW H  N  W+  + +  SA+  E N +I  +
Sbjct: 547 MVDLNKMYQAILALGRCWLHANNTAWSW-NFDTRSAIIAEYNAHINNL 593
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2591**

A DNA sequence (GASx699) was identified in *S.pyogenes* <SEQ ID 7681> which encodes the amino acid sequence <SEQ ID 7682>. Analysis of this protein sequence reveals the following:

```
Possible site: 36

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3323(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2592**

A DNA sequence (GASx701) was identified in *S.pyogenes* <SEQ ID 7683> which encodes the amino acid sequence <SEQ ID 7684>. Analysis of this protein sequence reveals the following:

```
Possible site: 20

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1017(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2593**

A DNA sequence (GASx702) was identified in *S.pyogenes* <SEQ ID 7685> which encodes the amino acid sequence <SEQ ID 7686>. Analysis of this protein sequence reveals the following:

```
Possible site: 27

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -3.03   Transmembrane    2 -  18 (   1 -  23)

----- Final Results -----
               bacterial membrane --- Certainty=0.2211(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2594**

A DNA sequence (GASx703) was identified in *S.pyogenes* <SEQ ID 7687> which encodes the amino acid sequence <SEQ ID 7688>. Analysis of this protein sequence reveals the following:

```
Possible site: 25

>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -3.45   Transmembrane   36 -  52 (  36 -  55)

----- Final Results -----
               bacterial membrane --- Certainty=0.2381(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC39287 GB:AF115103 orf87 gp [Streptococcus thermophilus
          bacteriophage Sfi21]
 Identities = 43/73 (58%), Positives = 61/73 (82%)

Query: 1  MINLKLRLQNKVTLMAILGAIFLLAQQLGIKLPSNIADIANTAVTLLVLLGVVTDPTTKG 60
          MIN KLRLQNK TL+A++ A+FL+ QQ G+ +P+NI +  NT V +LV+LG++TDPTTKG
Sbjct: 8  MINFKLRLQNKATLVALISAVFLMLQQFGLHVPNNIQEGINTLVGILVILGIITDPTTKG 67

Query: 61 LSDSEQALTYHEP 73
          ++DSE+AL+Y +P
Sbjct: 68 IADSERALSYIQP 80
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2595**

A DNA sequence (GASx707R) was identified in *S.pyogenes* <SEQ ID 7689> which encodes the amino acid sequence <SEQ ID 7690>. Analysis of this protein sequence reveals the following:

```
Possible site: 22

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-10.35   Transmembrane    9 -  25 (   1 -  27)

----- Final Results -----
               bacterial membrane --- Certainty=0.5140(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2596**

A DNA sequence (GASx714R) was identified in *S.pyogenes* <SEQ ID 7691> which encodes the amino acid sequence <SEQ ID 7692>. Analysis of this protein sequence reveals the following:

```
Possible site: 26

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1401(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2597**

A DNA sequence (GASx715) was identified in *S.pyogenes* <SEQ ID 7693> which encodes the amino acid sequence <SEQ ID 7694>. Analysis of this protein sequence reveals the following:

```
Possible site: 20

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0417(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2598**

A DNA sequence (GASx726) was identified in *S.pyogenes* <SEQ ID 7695> which encodes the amino acid sequence <SEQ ID 7696>. Analysis of this protein sequence reveals the following:

```
Possible site: 33

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.17   Transmembrane   18 -  34 (  18 -  35)

----- Final Results -----
               bacterial membrane --- Certainty=0.1468(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2599**

A DNA sequence (GASx728R) was identified in *S.pyogenes* <SEQ ID 7697> which encodes the amino acid sequence <SEQ ID 7698>. Analysis of this protein sequence reveals the following:

```
Possible site: 29

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1795(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF61314 GB:U96166 unknown [Streptococcus cristatus]
 Identities = 149/194 (76%), Positives = 162/194 (82%)

Query: 1   LSAIIRQSTSKRISDKRGIYLVEKLVSLAKQSYFTVTKTSPMIEEVRYYAKELLRLSERR 60
            L  IIRQSTSKRIS+KR  YL +KL+ LAKQS+  V KTSPM+EEVRYYA+ELLRLSERR
Sbjct: 56  LYEIIRQSTSKRISEKRIAYLTDKLIKLAKQSFCAVKKTSPMLEEVRYYAQELLRLSERR 115

Query: 61  QAIFDKMVASAQPLPEDKILRSIPSIVETTATSIIGELGAIRRFQSANQINAFIGIDFRH 120
            Q + + MVA AQPLPE  ILRSIP I ETTATSIIGELG I RFQS NQ NAFIGID RH
Sbjct: 116 QVVLNDMVALAQPLPEYDILRSIPGIAETTATSIIGELGDIHRFQSTNQFNAFIGIDLRH 175

Query: 121 YESGNYLAQEHITKRGNPYAPKILFKCIHDIAFASHTNPCHIADFYEKRKRQSQTASTKP 180
            YES N+LA+EHITKRGNPYA KILFKCIH+IA ASHTNPCHIADFYEKRKRQS  ASTKP
Sbjct: 176 YESRNFLAKEHITKRGNPYARKILFKCIHNIASASHTNPCHIADFYEKRKRQSTIASTKP 235

Query: 181 HTIASRHCLVRQCF 194
             TIAS H L+R  +
Sbjct: 236 LTIASIHRLIRTMY 249
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2600**

A DNA sequence (GASx729R) was identified in *S.pyogenes* <SEQ ID 7699> which encodes the amino acid sequence <SEQ ID 7700>. Analysis of this protein sequence reveals the following:

```
Possible site: 28

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2363(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2601**

A DNA sequence (GASx730R) was identified in *S.pyogenes* <SEQ ID 7701> which encodes the amino acid sequence <SEQ ID 7702>. Analysis of this protein sequence reveals the following:

```
Possible site: 25

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2602**

A DNA sequence (GASx734) was identified in *S.pyogenes* <SEQ ID 7703> which encodes the amino acid sequence <SEQ ID 7704>. Analysis of this protein sequence reveals the following:

```
Possible site: 52

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4001(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2603**

A DNA sequence (GASx735) was identified in *S.pyogenes* <SEQ ID 7705> which encodes the amino acid sequence <SEQ ID 7706>. Analysis of this protein sequence reveals the following:

```
Possible site: 55

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -3.66   Transmembrane  276 - 292 ( 274 - 292)

----- Final Results -----
              bacterial membrane --- Certainty=0.2466(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2604**

A DNA sequence (GASx736) was identified in *S.pyogenes* <SEQ ID 7707> which encodes the amino acid sequence <SEQ ID 7708>. Analysis of this protein sequence reveals the following:

```
Possible site: 33

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.3998(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2605**

A DNA sequence (GASx737) was identified in *S.pyogenes* <SEQ ID 7709> which encodes the amino acid sequence <SEQ ID 7710>. Analysis of this protein sequence reveals the following:

```
Possible site: 60

>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-12.74   Transmembrane   77 -  93 (  69 -  99)
   INTEGRAL    Likelihood = -4.14   Transmembrane  152 - 168 ( 151 - 170)
   INTEGRAL    Likelihood = -1.17   Transmembrane  196 - 212 ( 194 - 212)

----- Final Results -----
              bacterial membrane --- Certainty=0.6095(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2606**

A DNA sequence (GASx738) was identified in *S.pyogenes* <SEQ ID 7711> which encodes the amino acid sequence <SEQ ID 7712>. Analysis of this protein sequence reveals the following:

```
Possible site: 37

>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-13.16   Transmembrane   44 -  60 (  39 -  71)
   INTEGRAL    Likelihood =-10.24   Transmembrane   94 - 110 (  81 - 114)
   INTEGRAL    Likelihood = -7.64   Transmembrane  185 - 201 ( 179 - 207)
   INTEGRAL    Likelihood = -7.48   Transmembrane  132 - 148 ( 130 - 158)
   INTEGRAL    Likelihood = -2.76   Transmembrane  208 - 224 ( 204 - 225)
   INTEGRAL    Likelihood = -0.06   Transmembrane  153 - 169 ( 152 - 169)

----- Final Results -----
               bacterial membrane --- Certainty=0.6265(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2607**

A DNA sequence (GASx742) was identified in *S.pyogenes* <SEQ ID 7713> which encodes the amino acid sequence <SEQ ID 7714>. Analysis of this protein sequence reveals the following:

```
Possible site: 22

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -7.80   Transmembrane  887 - 903 ( 882 - 906)
   INTEGRAL    Likelihood = -4.88   Transmembrane    6 -  22 (   5 -  23)

----- Final Results -----
               bacterial membrane --- Certainty=0.4121(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

LPXTG motif: 877-881
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB46409 GB:AL096743 putative large secreted protein
           [Streptomyces coelicolor A3(2)]
 Identities = 231/599 (38%), Positives = 329/599 (54%), Gaps = 43/599 (7%)

Query: 278 TSSNSDASSRNIVKIGEIQGASHTSPLLKKAVTVEQVVVTYL---DDSTHFYVQDLNGDG 334
           T +++ +++   V+I ++QG++  SP    + VT    +VT +      S  F++QD   D
Sbjct: 28  TPAHAASAAAGPVRIHDVQGSTRLSPYAGEQVTDVAGIVTGVRGYGSSKGFWMQDPLPDA 87
```

```
Query: 335 DLATSDGIRVFAKNA-KVQVGDVLTISGEVEEFFGRGYEERKQTDLTITQIVAKAVTK-T 392
           D ATS+G+ VF   A +V VGD +T+SG V E+   G     Q+   +T+I    VT  +
Sbjct: 88  DPATSEGVFVFTSRAPEVAVGDAVTVSGTVSEYVPGGTSSGNQS---LTEITRPTVTVVS 144

Query: 393 GTAQVPSPLVLGKDRIAPANIIDNDGLR-------VFDPEEDAIDYWESMEGMLVAVDDA 445
           G   +P+   +    +  A   + DG            P   A+DY+ES+EGM V V DA
Sbjct: 145 GGNAIPAATTVSARSVPRAYAPEGDGAANGSVNALPLRPGTYALDYYESLEGMNVRVADA 204

Query: 446 KILGPMKN-KEIYVLPGSSTRPLNNSGGVLLPANSYNTDVIPVLFKKGKQI----IKAGD 500
           +++G      E++V       P    G V    +  NT  + +    GK         GD
Sbjct: 205 RVVGASDPYTELWVTVKPWENPNRRGGTVYGSYDDQNTGRLQIQ-SLGKPADFPAADVGD 263

Query: 501 SYKGRLAGPVSYS-YGNYKVFVDDSKNMPSLMDGHLKPEKTNLQKDLSKLSIASYNIENF 559
           +  G  AGP+ Y+ YG Y +   +   + +L  G  + E T  Q    +L++A+YN+EN
Sbjct: 264 TLAGTTAGPLDYNQYGGYTLVASE---IGALESGGTERESTRRQS-ARELAVATYNVENL 319

Query: 560 SANPSSTKDEKVKRIAESFIHDLNAPDIIGLIEVQDNNGPTDDGTTDATQSAQRLIDAIK 619
             +PS   D+     AE+ +H L +PDI+ L E+QDNNG TDDGT  A  +  RLIDAI
Sbjct: 320 --DPS---DDTFTAHAETIVHRLKSPDIVSLEEIQDNNGATDDGTVAADATVGRLIDAIV 374

Query: 620 KLGGPTYRYVDIAPENNVDGGQPGGNIRTGFLYQPERVSLSDKPKGGARDA--LTWVNGE 677
             GGP Y +  I P +  DGGQPGGNIR  FL+ PERVS +D+  G A  A  +  V G+
Sbjct: 375 AAGGPRYDWRGIDPVDKADGGQPGGNIRQAFLFNPERVSFTDRAGGDATTATGVRKVRGK 434

Query: 678 --LNLSVGRIDPTNAAWKDVRKSLAAEFIFQGRKVVVVANHLNSKRGDNALYGCVQPVTF 735
             L  S GR+DP N AW+D RK LA EF+F+GR V VVANH NSK GD  L    QP +
Sbjct: 435 AALTHSPGRVDPANEAWEDSRKPLAGEFVFRGRTVFVVANHFNSKGGDQGLTAQYQPPSR 494

Query: 736 KSEQRRHVLANMLAQFAKE--GAKHQANIVMLGDFNDFEFTKTIQLIE-EGDMVNLVSRH 792
            SE +RH  A ++  F KE   A+  A++V LGD NDFEF++T +++E +G + + V
Sbjct: 495 GSETQRHAQAKVVNTFVKEILAAQKNADVVALGDINDFEFSRTARILEGDGALWSAVKSL 554

Query: 793 DISDRYSYFHQGNNQTLDNILVSRHLL--DHYEFDMVHVNSPFMEAHGRASDHDPLLLQ 849
             S+RYSY +QGN+Q LD ILVS  +    H  +D VHVN+ F   H + SDHDP +L+
Sbjct: 555 PRSERYSYVYQGNSQVLDQILVSPSVRRGGHLSYDSVHVNAEF---HDQISDHDPQVLR 610
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2608**

A DNA sequence (GASx743) was identified in *S.pyogenes* <SEQ ID 7715> which encodes the amino acid sequence <SEQ ID 7716>. Analysis of this protein sequence reveals the following:

```
Possible site: 22

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2437(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2609**

A DNA sequence (GASx756) was identified in *S.pyogenes* <SEQ ID 7717> which encodes the amino acid sequence <SEQ ID 7718>. Analysis of this protein sequence reveals the following:

```
Possible site: 18

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -4.30    Transmembrane    10 -  26 (   8 -  27)
   INTEGRAL    Likelihood = -3.08    Transmembrane    51 -  67 (  50 -  67)

----- Final Results -----
               bacterial membrane --- Certainty=0.2720(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2610**

A repeated DNA sequence (GASx758) was identified in *S.pyogenes* <SEQ ID 7719> which encodes the amino acid sequence <SEQ ID 7720>. Analysis of this protein sequence reveals the following:

```
Possible site: 22

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA38133 GB:X54225 7 kDa protein [Streptococcus pneumoniae]
 Identities = 31/61 (50%), Positives = 41/61 (66%)

Query: 1  MTNGLKYVLEQMLLLFIIAALACLFLAIGLMIGYSFMGDGQSPWHILSMDKWAELVNKFT 60
          M     YV++++LL+ I+  L  L L IGLM+GY  +G GQ PW ILS  KW EL++KFT
Sbjct: 3  MNKKSSYVVKRLLLVIIVLILGTLALGIGLMVGYGILGKGQDPWAILSPAKWQELIHKFT 62

Query: 61 G 61
          G
Sbjct: 63 G 63
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2611**

A DNA sequence (GASx764) was identified in *S.pyogenes* <SEQ ID 7721> which encodes the amino acid sequence <SEQ ID 7722>. Analysis of this protein sequence reveals the following:

```
Possible site: 58

>>> Seems to have no N-terminal signal sequence
```

```
      INTEGRAL    Likelihood = -3.98   Transmembrane   47 -  63 (  46 -  67)

   ----- Final Results -----
                  bacterial membrane --- Certainty=0.2593(Affirmative) < succ>
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related sequence was also identified in GAS <SEQ ID 9149> which encodes the amino acid sequence <SEQ ID 9150>. Analysis of this protein sequence reveals the following:

```
     Possible site: 53
   >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -3.98   Transmembrane   35 -  51 (  34 -  55)

   ----- Final Results -----
                  bacterial membrane --- Certainty=0.2593(Affirmative) < succ>
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2612**

A DNA sequence (GASx783) was identified in *S.pyogenes* <SEQ ID 7723> which encodes the amino acid sequence <SEQ ID 7724>. Analysis of this protein sequence reveals the following:

```
   Possible site: 43

   >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood =-13.16   Transmembrane  142 - 158 ( 132 - 167)
      INTEGRAL    Likelihood =-12.26   Transmembrane  113 - 129 ( 101 - 140)
      INTEGRAL    Likelihood =-10.24   Transmembrane  238 - 254 ( 233 - 260)
      INTEGRAL    Likelihood = -2.76   Transmembrane   34 -  50 (  34 -  51)

   ----- Final Results -----
                  bacterial membrane --- Certainty=0.6265(Affirmative) < succ>
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAA32091 GB:AB010970 ABC-transporter [Streptococcus mutans]
 Identities = 173/269 (64%), Positives = 214/269 (79%), Gaps = 2/269 (0%)

Query: 1   MNFLTKKNRILLREMVKTDFKLRYQGSAIGYLWSILKPLMMFTIMYLVFIRFLRLGGNVP 60
            M+F ++KNRILL+E++KTDFKLRYQGSAIGYLWSILKPLM+F IMY+VF+RFL LGG+VP
Sbjct: 1   MDFFSRKNRILLKELIKTDFKLRYQGSAIGYLWSILKPLMLFAIMYIVFVRFLPLGGDVP 60

Query: 61  HFPVALLLANVIWSFFSEATSMGMVSIVSRGDLLRKLNFSKHIIVFSAVLGALINFLINL 120
            H+PVALLL NVIW+FF E T MGMVS+V+RGDLLRKLNFSK  IVFSAV GA INF IN+
Sbjct: 61  HWPVALLLGNVIWTFFQETTMMGMVSVVTRGDLLRKLNFSKQTIVFSAVSGAAINFGINV 120

Query: 121 VVVLIFALINGVTIS--GYAYLSLFLFIELVVLVLGIALLLSNVFVYYRDLAQVWEVLLQ 178
            +VVLIFAL+NGVT +     +L + LF+EL++   GIA +LS ++V YRD+  VWEV+LQ
Sbjct: 121 IVVLIFALLNGVTFTFRWNLFLLIPLFLELLLFSTGIAFILSTLYVRYRDIGPVWEVILQ 180

Query: 179 AGMYATPIIYPITFVLDSHPLAAKLLMLNPVAQMIQDFRYLLIDRANVTIWQMSTNWFYI 238
```

```
            G Y TPIIY +T++     + AKLL+L+P+AQ+IQD R++LID ANVTIWQM  +
Sbjct: 181 GGFYGTPIIYSLTYIATRSVVGAKLLLLSPIAQIIQDMRHILIDPANVTIWQMINHKSIA 240

Query: 239 VIPYLVPFVILFIGIFVFKKNADRFAEII 267
           VIPYLVP  +  IG  VF  NA +FAEII
Sbjct: 241 VIPYLVPIFVFIIGFLVFNYNAKKFAEII 269
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2613**

A DNA sequence (GASx786) was identified in *S.pyogenes* <SEQ ID 7725> which encodes the amino acid sequence <SEQ ID 7726>. Analysis of this protein sequence reveals the following:

```
Possible site: 32

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3828(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAA32094 GB:AB010970 rgpFc [Streptococcus mutans]
 Identities = 381/582 (65%), Positives = 475/582 (81%), Gaps = 1/582 (0%)

Query: 1   MNRILLYVHFNKYNKISAHVYYQLEQMRSLFSKIVFISNSKVSHEDLKRLKNHCLIDEFL 60
           M R+LLYVHFNKYN++S+HV YQL QMRSLFSK++FISNS+V+  D+K L+   LID+F+
Sbjct: 1   MKRLLLYVHFNKYNRVSSHVVYQLTQMRSLFSKVIFISNSQVADADVKMLREKHLIDDFI 60

Query: 61  QRKNKGFDFSAWHDGLIIMGFDKLEEFDSLTIMNDTCFGPIWEMAPYFENFEEKETVDFW 120
           QR+N GFDF+AW DG++ +GFD+L  +DS+T MNDTCFGP+WEM   ++ FE K TVDFW
Sbjct: 61  QRQNSGFDFAAWRDGMVFVGFDELVTYDSVTTMNDTCFGPLWEMYSIYQEFETKTTVDFW 120

Query: 121 GITNNRGTKAFKEHVQSYFMTFKNQVIQNKVFQQFWQSIIEYENVQEVIQHYETQLTSIL 180
           G+TNNR TK+F+EH+QSYF++FK  V+++  F+ FW++I EY++VQ+VI  YET++T+ L
Sbjct: 121 GLTNNRATKSFREHIQSYFISFKASVLRSTAFRDFWENIKEYQDVQKVIDQYETKVTTTL 180

Query: 181 LNEGFSYQTVFDTRKAESSFMPHPDFSYYNPTAILKHHVPFIKVKAIDANQHIAPYLLNL 240
           L+ GF Y  VFDT K ++S M H DFSYYNPTAIL H VPFIKVKAID NQHI PYLLN
Sbjct: 181 LDAGFQYDVVFDTTKEDASHMLHADFSYYNPTAILNHRVPFIKVKAIDNNQHITPYLLND 240

Query: 241 IRETTNYPIDLIVSHMSQISLPDTKYLLSQKYLNCQRLAKQTCQKVAVHLHVFYVDLLDE 300
           I++ + YPIDLIVSHMS+I+ PD  YLL  KY+  +       QKVAVHLHVFYVDLL+E
Sbjct: 241 IQKNSTYPIDLIVSHMSEINYPDFSYLLGHKYVKKRERVDLKNQKVAVHLHVFYVDLLEE 300

Query: 301 FLTAFENWNFHYDLFITTDSDIKRKEIKEILQRKGKTADIRVTGNRGRDIYPMLLLKDKL 360
           FLTAF+ ++F YDLFITTDSD K+ EI+EIL   G+ A + VTGN GRD+ PML LK+ L
Sbjct: 301 FLTAFKQFHFSYDLFITTDSDDKKAEIEEILSANGQEAQVFVTGNIGRDVLPMLKLKNYL 360

Query: 361 SQYDYIGHFHTKKSKEADFWAGESWRKELIDMLVKPADSILSAFETD-DIGIIIADIPSF 419
           S YD++GHFHTKKSKEADFWAG+SWR+ELIDMLVKPAD+IL+  + +  IG++IAD+P+F
Sbjct: 361 SAYDFVGHFHTKKSKEADFWAGQSWREELIDMLVKPADNILAQLQQNPKIGLVIADMPTF 420

Query: 420 FRFNKIVNAWNEHLIAQEMMSLWRKMDVKKQIDFQAMDTFVMSYGTFVWFKYDALKSLFD 479
           FR+NKIV+AWNEHLIA EM +LW+KM + K+IDF A  TFVMSYGTFVWFKYDALK LFD
Sbjct: 421 FRYNKIVDAWNEHLIAPEMNTLWQKMGMTKKIDFNAFHTFVMSYGTFVWFKYDALKPLFD 480

Query: 480 LELTQNDIPSEPLPQNSILHAIERLLVYIAWGDSYDFRIVKNPYELTPFIDNKLLNLRED 539
           L LT +D+P EPLPQNSILHAIERLL+YIAW + YDFRI KNP +LTPFIDNKLLN R +
Sbjct: 481 LNLTDDDVPEEPLPQNSILHAIERLLIYIAWNEHYDFRISKNPVDLTPFIDNKLLNERGN 540
```

```
Query: 540 EGAHTYVNFNQMGGIKGALKYIIVGPAKAMKYIFLRLMEKLK 581
              +T+V+FN MGGIKGA KYI +GPA+A+KYI  R ++K+K
Sbjct: 541 SAPNTFVDFNYMGGIKGAFKYIFIGPARAVKYILKRSLQKIK 582
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2614**

A DNA sequence (GASx787) was identified in *S.pyogenes* <SEQ ID 7727> which encodes the amino acid sequence <SEQ ID 7728>. Analysis of this protein sequence reveals the following:

```
Possible site: 33

>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-15.66   Transmembrane  202 - 218 ( 191 - 224)
   INTEGRAL    Likelihood =-10.03   Transmembrane  340 - 356 ( 335 - 365)
   INTEGRAL    Likelihood = -9.08   Transmembrane  270 - 286 ( 263 - 289)
   INTEGRAL    Likelihood = -8.60   Transmembrane  124 - 140 ( 118 - 145)
   INTEGRAL    Likelihood = -4.94   Transmembrane  377 - 393 ( 375 - 395)
   INTEGRAL    Likelihood = -3.29   Transmembrane  291 - 307 ( 290 - 311)
   INTEGRAL    Likelihood = -2.87   Transmembrane  160 - 176 ( 159 - 180)
   INTEGRAL    Likelihood = -2.66   Transmembrane   50 -  66 (  48 -  66)
   INTEGRAL    Likelihood = -1.28   Transmembrane   77 -  93 (  76 -  93)
   INTEGRAL    Likelihood = -0.69   Transmembrane  229 - 245 ( 229 - 245)

----- Final Results -----
               bacterial membrane --- Certainty=0.7262(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAA32095 GB:AB010970 ORF7 [Streptococcus mutans]
 Identities = 374/775 (48%), Positives = 525/775 (67%), Gaps = 7/775 (0%)

Query: 53  VSFVGYIISLIGLSYYLSRQVSRQLFLKTSFIVISYLIVSYWVQITQHLNDKRFDIWSLT 112
            V  V Y++S++GLS+YLS+ + +  F++       Y+++SY++ +T+ LN++ F IW L
Sbjct: 30  VCLVIYVLSILGLSFYLSKNLKKTFFIELLLGYGLYIVISYFLAVTRELNNESFKIWDLA 89

Query: 113 KNQFYQFQALPSLLIILV----MATLIKILAAYFAIEKDRFGLL-GYQGNTFSVALILAV 167
            KN F+Q   LP+L++I+     +  LI++      + +    LL  +    F +   ++
Sbjct: 90  KNHFFQPYFLPTLVLIIACTFALNYLIRVKMKRSHLSRKMTLLLENFSETEFLLTGLIVS 149

Query: 168 VPINDIHLLKLISSRFSELVTAGNSQIALLKISGLLIVLLVIFATIIYVVLNALKHLKSN 227
              ++D   +KL+            +  +LL +  LL   L++F+ I+    NA + +K N
Sbjct: 150 FILSDTLYVKLLQESLRAYYHKPLAYESLLFLYTLLT--LILFSVIVEACFNAYRSIKLN 207

Query: 228 KPSFSVAATTSLFLALVFNYTFQYGVKGDEALLGYYVFPGATLFQIVAITLVALLAYVIT 287
            +P+ S+A  +SL  A +FNY FQYG+K D  LLG Y+ PGAT +QI+ +T      Y+I
Sbjct: 208 RPNLSLAFVSSLLFATIFNYAFQYGLKNDADLLGKYIVPGATAYQILVLTAAGFFLYLII 267

Query: 288 NRYWPTTFFLLILGTIISVVNDLKESMRSEPLLVTDFVWLQELGLVTSFVKKSVIVEMVV 347
            NRY   TF ++ILG+II+VVN LK  MR+EPLLVTDF W+  + L+   V  ++I   ++
Sbjct: 268 NRYLLVTFLIVILGSIITVVNVLKVGMRNEPLLVTDFAWVTNIRLLARSVNANIIFSTLL 327

Query: 348 GLAICIVVAWYLHGRVLAGKLFMSPVKRASAVLGLFIVSCSMLIPFSYEKEGKILSGLPI 407
             LA  I++  +L  R+L GK+  +   +   +  + ++  S+ I F  EK  KI++G+P+
Sbjct: 328 ILAALILLYLFLRKRLLQGKITENYRLKVGLISSICLLGFSIFIIFRNEKGSKIVNGIPV 387

Query: 408 ISALNNDNDINWLGFSTNARYKSLAYVWTRQVTKKIMEKPTNYSQETIASIAQKYQKLAE 467
            IS +NN  DI + GF +NA YKSL YVWT+QVTK IM+KP++YS+E I  +A+KY  +A
Sbjct: 388 ISQVNNWVDIGYQGFYSNASYKSLMYVWTKQVTKSIMDKPSDYSKERILKLAKKYNNVAN 447
```

```
Query: 468 DINKDRKNNIADQTVIYLLSESLSDPDRVSNVTVSHDVLPNIKAIKNSTTAGLMQSDSYG 527
            INK R  NI++QTVIY+LSES SDPDRV  V +S DV+PNIK IK  TT+GLM SD YG
Sbjct: 448 KINKVRTENISNQTVIYILSESFSDPDRVQGVNLSRDVIPNIKQIKEKTTSGLMHSDGYG 507

Query: 528 GGTANMEFQTLTSLPFYNFSSSVSVLYSEVFPKMAKPHTISEFYQGKNRIAMHPASANNF 587
           GGTANMEFQ+LT LP+YNF+SSVS LY+EV P M+   +IS  ++ KNR+ +HP+SA+N+
Sbjct: 508 GGTANMEFQSLTGLPYYNFNSSVSTLYTEVVPDMSVFPSISNQFKSKNRVVIHPSSASNY 567

Query: 588 NRKTVYSNLGFSKFLALSGSKDKFKNIENVGLLTSDKTVYNNILSLINPSESQFFSVITM 647
           +RK VY  L F  F+A SG+ DK  + E VGL  SDKT Y NIL  INPS+SQFFSV+TM
Sbjct: 568 SRKYVYDKLKFPTFVASSGTSDKITHSEKVGLNVSDKTTYQNILDKINPSQSQFFSVMTM 627

Query: 648 QNHIPWSSDYPEEIVAEGKNFTEEENHNLTSYARLLSFTDKETRAFLEKLTQINKPITVV 707
           QNH+PW+SD P ++VA GK +T++EN +L+SYARLL++TDKET+ FL +L+Q+   +TVV
Sbjct: 628 QNHVPWASDEPSDVVATGKGYTKDENGSLSSYARLLTYTDKETKDFLAQLSQLKHKVTVV 687

Query: 708 FYGDHLPGLYPDSAFNKHIENKYLTDYFIWSNGTNEKKNHPLINSSDFTAALFEHTDSKV 767
           FYGDHLPGLYP+SAF K  +++Y TDYFIWSN   +  NH  +NSSDFTA L EHT+SKV
Sbjct: 688 FYGDHLPGLYPESAFKKDPDSQYQTDYFIWSNYNTKTLNHSYVNSSDFTAELLEHTNSKV 747

Query: 768 SPYYALLTEVLNKASVDKSPDSPEVKAIQNDLKNIQYDVTIGKGYLLKHKTFFKI 822
           SPYYALLTEVL+  +V     + E K I NDLK IQYD+T+GKGY+  +K FF I
Sbjct: 748 SPYYALLTEVLDNTTVGHGKLTKEQKEIANDLKLIQYDITVGKGYIRNYKGFFDI 802
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2615**

A DNA sequence (GASx789R) was identified in *S.pyogenes* <SEQ ID 7729> which encodes the amino acid sequence <SEQ ID 7730>. Analysis of this protein sequence reveals the following:

```
Possible site: 13

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.06    Transmembrane   42 -  58 (  42 -  58)

----- Final Results -----
               bacterial membrane --- Certainty=0.1426(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2616**

A DNA sequence (GASx790) was identified in *S.pyogenes* <SEQ ID 7731> which encodes the amino acid sequence <SEQ ID 7732>. Analysis of this protein sequence reveals the following:

```
Possible site: 24

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

## Example 2617

A DNA sequence (GASx791) was identified in *S.pyogenes* <SEQ ID 7733> which encodes the amino acid sequence <SEQ ID 7734>. Analysis of this protein sequence reveals the following:

```
Possible site: 48

>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-12.42   Transmembrane  166 - 182 ( 157 - 188)
   INTEGRAL    Likelihood = -7.32   Transmembrane   85 - 101 (  79 - 104)
   INTEGRAL    Likelihood = -6.90   Transmembrane  397 - 413 ( 386 - 417)
   INTEGRAL    Likelihood = -6.05   Transmembrane  253 - 269 ( 252 - 273)
   INTEGRAL    Likelihood = -5.26   Transmembrane  301 - 317 ( 293 - 325)
   INTEGRAL    Likelihood = -3.35   Transmembrane  363 - 379 ( 362 - 379)
   INTEGRAL    Likelihood = -3.24   Transmembrane  335 - 351 ( 335 - 351)

----- Final Results -----
              bacterial membrane --- Certainty=0.5967(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae.*

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAA64645 GB:U10927 CapF [Staphylococcus aureus]
 Identities = 97/419 (23%), Positives = 186/419 (44%), Gaps = 40/419 (9%)

Query: 12  FLWNMLGSLSTAVISVILLMVVTRLLTSADSDIYAFAYSFANMMVVVGLFQVRNYQATDI 71
           F +  + ++ +A+   ++L+V+ RL T  D   Y +A      + +    ++R+   T
Sbjct: 5   FNYMFVANILSALCKFLILLVIVRLGTPEDVGRYNYALVITAPIFLFISLKIRSVIVT-- 62

Query: 72  NEKYSFSQYLVARLMTCLLMLAITVIYLTLTKTDSYKSTIVFLVCFYRSTDAFSDLYQGM 131
           N+KYS ++Y+ A L   ++ L    I++ +        T + +V   +  +   ++  G+
Sbjct: 63  NDKYSPNEYISAILSLNIITLIFVAIFVYVLGNGDL--TTILIVSLIKLFENIKEVPYGI 120

Query: 132 FQQHERLDIAGKSLAYRNTLIFMVYTAIILYSKNLTLALVAVCIVSLVFIMYYDIGHSKK 191
           +Q++E L + G S+   N L  +++  I  +S NL +AL+ + I  +      D  +  K
Sbjct: 121 YQKNESLKLLGISMGIYNILSLILFYIIYSFSHNLNMALLFLVISCIFSFAIIDRWYLSK 180

Query: 192 FQKLMFSELLSNISFQNSLKLLKESF----PLFLNGFLIIYIYTQPKYAIELMTTLGEVA 247
           +  +        + + N++   KE F    PL  +  L       P+  +E +   G+
Sbjct: 181 YYNI-------KLHYNNNIAKFKEIFILTIPLAFSSALGSLNTGIPRIVLENL--FGKYT 231

Query: 248 LGS-QTIFNILFMPAFVMNLLILFFRPHITQMAIALIRGQIK-EFNKIQVQLFAYLGVF- 304
           LG   TI  +L +     N +   F P + +    L + + K EF K+  ++  ++G+F
Sbjct: 232 LGIFSTIAYVLVIGGLFANSISQVFLPKLRK----LYKDEKKIEFEKLTRKM-VFIGIFI 286

Query: 305 SLIALVGSGLFGIPFLSILYG-----TNLTDYWVDF-MLIMLGGSIGSFATVIDNILTAM 358
            + +++ S   G   LS+L+G      N+    + F +L +L G       +
Sbjct: 287 GMCSVILSLFLGEALLSLLFGKEYGENNIILIILSFGLLFILSGIFLGTTIIATGKYNVN 346

Query: 359 RKQQLLLIPYTGGFLISLLITNLFVMKYHILGAALSFLITMLVWLGLSIMIYLFIMNRF 417
            K  L+L+     F I L+ + L + KY +LGAAL+  I+  V L   I  Y F    F
Sbjct: 347 YKISLILL-----FCI-LIFSFLLIPKYSLLGAALTITISQFVAL---ISYYYFYKRIF 396
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2618

A DNA sequence (GASx792) was identified in *S.pyogenes* <SEQ ID 7735> which encodes the amino acid sequence <SEQ ID 7736>. Analysis of this protein sequence reveals the following:

```
Possible site: 36

>>> Seems to have no N-terminal signal sequence
     INTEGRAL    Likelihood =-10.03    Transmembrane   64 -  80 (  60 -  84)
     INTEGRAL    Likelihood = -9.66    Transmembrane   43 -  59 (  37 -  63)
     INTEGRAL    Likelihood = -8.70    Transmembrane  232 - 248 ( 229 - 251)
     INTEGRAL    Likelihood = -8.28    Transmembrane  410 - 426 ( 402 - 432)
     INTEGRAL    Likelihood = -6.21    Transmembrane  298 - 314 ( 296 - 322)
     INTEGRAL    Likelihood = -6.21    Transmembrane  478 - 494 ( 471 - 496)
     INTEGRAL    Likelihood = -5.04    Transmembrane  265 - 281 ( 256 - 288)
     INTEGRAL    Likelihood = -3.29    Transmembrane  380 - 396 ( 378 - 397)
     INTEGRAL    Likelihood = -2.92    Transmembrane  210 - 226 ( 209 - 227)
     INTEGRAL    Likelihood = -2.60    Transmembrane  187 - 203 ( 187 - 204)
     INTEGRAL    Likelihood = -2.50    Transmembrane  442 - 458 ( 439 - 458)
     INTEGRAL    Likelihood = -1.65    Transmembrane   18 -  34 (  18 -  35)
     INTEGRAL    Likelihood = -1.38    Transmembrane  165 - 181 ( 165 - 181)

----- Final Results -----
                  bacterial membrane --- Certainty=0.5012(Affirmative) < succ>
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAA19642 GB:AB002668 unnamed protein product [Actinobacillus
           actinomycetemcomitans]
 Identities = 116/459 (25%), Positives = 207/459 (44%), Gaps = 60/459 (13%)

Query: 69  FILVFGTISAIISPINDIPDEYVHYSRTVYISEGDINLTNNNKKLRISKDVDKLI----- 123
           FIL F  I  II+P    PDE+ H+ R   IS G I  ++  K    + K + K++
Sbjct: 16  FILTF-IIGVIITPPYQSPDEFYHFQRGYAISNGQIIPSSTEK---LDKAMMKMLSIYEG 71

Query: 124 ----KQSGKTFITSNLKATKHSTREYSYPYIKGTNAYYSFSYIPQALGILVGNALDLPIL 179
                ++  T    N        +EY       TN Y+   Y+PQALG  +G+ LDL +
Sbjct: 72  IPYRSENKVTHFLENEAQNVAWEKEYILDESANTNVYFPLIYLPQALGSFLGSTLDLSLY 131

Query: 180 LTYYFGRLCN-LISYAMLAFIAIKLSGSFKQVIAVVTLLPMNIYLAASFNQDGFAIGLVL 238
             YY  ++   L+S A+L F +++   S    + ++  LPM ++    S N D      ++
Sbjct: 132 NMYYLAKIFTLLVSIAILYFASVQYRLSIP--VLLILSLPMTMFQMGSTNPDS-----II 184

Query: 239 VTIGLFI-NLLSSKDKSNYNTKFFLYLVLCGLL------VLSKFTYFLLVCLPLFIPNEK 291
            ++ +FI +LL+    SNYN   F +   C LL      V  KF  +L+  LP FI   +
Sbjct: 185 FSLSVFIGSLLARGLDSNYN---FTHKDFCKLLFSIFLCVTVKFNMLVLLLLPFFISKRR 241

Query: 292 FGKNTKLVILKKLGGLLLIFLFAAMWFRLYGQVKTPYVADFLKEV----NVSQQVKNMLE 347
             ++  +  +  +   +L  + A          K   + +F   +    ++ +  KN L
Sbjct: 242 EIRHGSMYSIFIIILSILWIVLAMKLTEAQSHFKEGALHNFSYYIFHMDDLFEIFKNTLN 301

Query: 348 SPIVYSSIIIRHMVINLINMNNIFQFGA-LSYGITNLFPLYVCFFFFVYISNASKITINI 406
             + Y   ++R +   L  ++  F     L +G T+L       + F++I N  K+   I
Sbjct: 302 --LTYLKSLLRMFLGVLGWVDTKFTINEYLFFGSTSLLA-----YIFLFIHNLYKLKYVI 354

Query: 407 VEKM--GIIFVISAIIGATVLAMYLTWTPVGSSTVLGVQSRYLIGIIPLVLLLFSS---- 460
           V  +  G++F+ +  I      + +T+  +G++ ++GVQ RY    IP++L++FSS
Sbjct: 355 VSVLLVGVVFLFTHFI------LLITYNEIGTTQIVGVQGRY---FIPIMLIIFSSFILK 405
```

```
Query: 461 QQQKFKQIEDILSDKLAIHVSLLFILAMLM--STIFRYY 497
           + +K    + I    + +    LFI + +   + + RYY
Sbjct: 406 KSEKTSNNKTISKYFIIVPFLFLFISSFITINTLVSRYY 444
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2619**

A DNA sequence (GASx797) was identified in *S.pyogenes* <SEQ ID 7737> which encodes the amino acid sequence <SEQ ID 7738>. Analysis of this protein sequence reveals the following:

```
Possible site: 49

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1491(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC83961 GB:L47648 cytidine monophosphate kinase [Bacillus subtilis]
 Identities = 116/220 (52%), Positives = 156/220 (70%), Gaps = 1/220 (0%)

Query: 2   KAIKIAIDGPASSGKSTVAKIIAKNLGYTYLDTGAMYRSATYIALTHGYTGKEVALILEE 61
           K + IAIDGPA++GKSTVAKI+A+   Y Y+DTGAMYR+ TY AL       +   + E
Sbjct: 3   KKLSIAIDGPAAAGKSTVAKIVAEKKSYIYIDTGAMYRAITYAALQENVDLTDEEKLAEL 62

Query: 62  LEKNPIFFKKAKDGSQLVFLGDEDVTLAIRQNDVTNNVSWISALPEIREELVHQQRRIAQ 121
           L++  I     KDG Q VF+   DVT AIR ++++N VS  +    +REE+V +Q+++ +
Sbjct: 63  LKRTDIELITTKDG-QKVFVNGTDVTEAIRTDEISNQVSIAAKHRSVREEMVKRQQQLGE 121

Query: 122 AGGIIMDGRDIGTVVLPDAELKIFLVASVEERAERRYKENLEKGIESDFETLKEEIAARD 181
            GG++MDGRDIGT VLP+AE+KIFL+ASVEERA+RRY+EN++KG + ++ETL EEIA RD
Sbjct: 122 KGGVVMDGRDIGTHVLPNAEVKIFLLASVEERAKRRYEENVKKGFDVNYETLIEEIARRD 181

Query: 182 YKDSHRKVSPLKAAEDALIFDTTGVSIDGVVQFIQEKAEK 221
             DS R+VSPL+ AEDAL  DTT +SI  V   I E  E+
Sbjct: 182 KLDSEREVSPLRKAEDALEIDTTSLSIQEVADKILEAVEQ 221
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2620**

A DNA sequence (GASx799) was identified in *S.pyogenes* <SEQ ID 7739> which encodes the amino acid sequence <SEQ ID 7740>. Analysis of this protein sequence reveals the following:

```
Possible site: 29

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4324(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA34313 GB:X16188 ribosomal protein L35 (AA 1-66) [Bacillus
          stearothermophilus]
 Identities = 46/65 (70%), Positives = 51/65 (77%)

Query: 1  MPKQKTHRASAKRFKRTGSGGLKRFRAFTSHRFHGKTKKQRRHLRKAGLVSSGDFKRIKA 60
          MPK KTHR SAKRFK+T SG LKR  A+TSH F  KTKKQ+RHLRKA LVS GDFKRI+
Sbjct: 1  MPKMKTHRGSAKRFKKTASGKLKRGHAYTSHLFANKTKKQKRHLRKATLVSPGDFKRIRQ 60

Query: 61 MVTGL 65
          M+  L
Sbjct: 61 MLDNL 65
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2621**

A DNA sequence (GASx806R) was identified in *S.pyogenes* <SEQ ID 7741> which encodes the amino acid sequence <SEQ ID 7742>. Analysis of this protein sequence reveals the following:

```
Possible site: 16

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5361(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2622**

A DNA sequence (GASx809R) was identified in *S.pyogenes* <SEQ ID 7743> which encodes the amino acid sequence <SEQ ID 7744>. Analysis of this protein sequence reveals the following:

```
Possible site: 52

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.81    Transmembrane   33 -  49 (  28 -  53)

----- Final Results -----
               bacterial membrane --- Certainty=0.4524(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2623**

A DNA sequence (GASx814R) was identified in *S.pyogenes* <SEQ ID 7745> which encodes the amino acid sequence <SEQ ID 7746>. Analysis of this protein sequence reveals the following:

```
Possible site: 33

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0206(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2624**

A DNA sequence (GASx817) was identified in *S.pyogenes* <SEQ ID 7747> which encodes the amino acid sequence <SEQ ID 7748>. Analysis of this protein sequence reveals the following:

```
Possible site: 13

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -1.49   Transmembrane   16 -  32 (  15 -  32)

----- Final Results -----
               bacterial membrane --- Certainty=0.1595(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2625**

A DNA sequence (GASx820) was identified in *S.pyogenes* <SEQ ID 7749> which encodes the amino acid sequence <SEQ ID 7750>. Analysis of this protein sequence reveals the following:

```
Possible site: 31

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -7.11   Transmembrane   62 -  78 (  59 -  81)
   INTEGRAL    Likelihood = -6.00   Transmembrane  128 - 144 ( 123 - 147)
   INTEGRAL    Likelihood = -2.50   Transmembrane    5 -  21 (   3 -  26)

----- Final Results -----
               bacterial membrane --- Certainty=0.3845(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAA26653 GB:M83994 prolipoprotein signal peptidase
           [Staphylococcus aureus]
 Identities = 57/153 (37%), Positives = 96/153 (62%), Gaps = 6/153 (3%)

Query: 1   MKKRLFVLSLILL----VALDQLSKFWIVSHIALGEVKPFIPGIVSLTYLQNNGAAFSIL 56
           M K+ F+ + IL+    V  DQ++K+ I + + +G+    IP  +++T  +NNGAA+ IL
Sbjct: 1   MHKKYFIGTSILIAVFVVIFDQVTKYIIATTMKIGDSFEVIPHFLNITSHRNNGAAWGIL 60

Query: 57  QDQQWFFVVITVLVIGYAIYYLATHPHLNIWKQLALLLIISGGIGNFIDRLRLAYVIDMI 116
             +  FF +IT++++   +Y+       N++ Q+A+ L+ +G +GNFIDR+    V+D I
Sbjct: 61  SGKMTFFFIITIIILIALVYFFIKDAQYNLFMQVAISLLFAGALGNFIDRILTGEVVDFI 120

Query: 117 HLDF--VDFAIFNVADSYLTVGVILLLICLWKE 147
             +    DF IFN+ADS LT+GVIL++I L K+
Sbjct: 121 DTNIFGYDFPIFNIADSSLTIGVILIIIALLKD 153
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2626**

A DNA sequence (GASx822R) was identified in *S.pyogenes* <SEQ ID 7751> which encodes the amino acid sequence <SEQ ID 7752>. Analysis of this protein sequence reveals the following:

```
Possible site: 33

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2638(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2627**

A DNA sequence (GASx823R) was identified in *S.pyogenes* <SEQ ID 7753> which encodes the amino acid sequence <SEQ ID 7754>. Analysis of this protein sequence reveals the following:

```
Possible site: 45

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3452(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2628**

A DNA sequence (GASx828) was identified in *S.pyogenes* <SEQ ID 7755> which encodes the amino acid sequence <SEQ ID 7756>. Analysis of this protein sequence reveals the following:

```
Possible site: 21

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2629**

A DNA sequence (GASx836) was identified in *S.pyogenes* <SEQ ID 7757> which encodes the amino acid sequence <SEQ ID 7758>. Analysis of this protein sequence reveals the following:

```
Possible site: 18

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.4333(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2630**

A DNA sequence (GASx853R) was identified in *S.pyogenes* <SEQ ID 7759> which encodes the amino acid sequence <SEQ ID 7760>. Analysis of this protein sequence reveals the following:

```
Possible site: 14

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.4906(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2631**

A DNA sequence (GASx854R) was identified in *S.pyogenes* <SEQ ID 7761> which encodes the amino acid sequence <SEQ ID 7762>. Analysis of this protein sequence reveals the following:

```
Possible site: 43

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3989(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related sequence was also identified in GAS <SEQ ID 9147> which encodes the amino acid sequence <SEQ ID 9148>. Analysis of this protein sequence reveals the following:

```
Possible site: 42

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty= 0.399(Affirmative) < succ>
               bacterial membrane --- Certainty= 0.000(Not Clear) < succ>
                bacterial outside --- Certainty= 0.000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB59092 GB:M97157 pyrogenic exotoxin C [Streptococcus pyogenes]
 Identities = 39/67 (58%), Positives = 53/67 (78%)

Query: 1   LMESKEIYLTKSPYIRGSLEIHSKNRKHEKINLYDAKPNSTRSDVFKKYKDNKTINMKDF 60
           LM++ +IY   SPY+ G +EI +K+ KHE+I+L+D+    TRSD+F KYKDN+ INMK+F
Sbjct: 167 LMDNYKIYDATSPYVSGRIEIGTKDGKHEQIDLFDSPNEGTRSDIFAKYKDNRIINMKNF 226

Query: 61  SHFDIYL 67
           SHFDIYL
Sbjct: 227 SHFDIYL 233
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2632**

A DNA sequence (GASx855R) was identified in *S.pyogenes* <SEQ ID 7763> which encodes the amino acid sequence <SEQ ID 7764>. Analysis of this protein sequence reveals the following:

```
Possible site: 33

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2633**

A DNA sequence (GASx856) was identified in *S.pyogenes* <SEQ ID 7765> which encodes the amino acid sequence <SEQ ID 7766>. Analysis of this protein sequence reveals the following:

```
Possible site: 26

>>> Seems to have no N-terminal signal sequence

-----, Final Results -----
              bacterial cytoplasm --- Certainty=0.4145(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2634**

A DNA sequence (GASx862) was identified in *S.pyogenes* <SEQ ID 7767> which encodes the amino acid sequence <SEQ ID 7768>. Analysis of this protein sequence reveals the following:

```
Possible site: 19

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.6285(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2635**

A DNA sequence (GASx863) was identified in *S.pyogenes* <SEQ ID 7769> which encodes the amino acid sequence <SEQ ID 7770>. Analysis of this protein sequence reveals the following:

```
Possible site: 51

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2636**

A DNA sequence (GASx878) was identified in *S.pyogenes* <SEQ ID 7771> which encodes the amino acid sequence <SEQ ID 7772>. Analysis of this protein sequence reveals the following:

```
Possible site: 21

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2637**

A DNA sequence (GASx887R) was identified in *S.pyogenes* <SEQ ID 7773> which encodes the amino acid sequence <SEQ ID 7774>. Analysis of this protein sequence reveals the following:

```
Possible site: 20

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1911(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2638**

A DNA sequence (GASx910) was identified in *S.pyogenes* <SEQ ID 7775> which encodes the amino acid sequence <SEQ ID 7776>. Analysis of this protein sequence reveals the following:

```
Possible site: 20

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4511(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```

```
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2639**

A DNA sequence (GASx911) was identified in *S.pyogenes* <SEQ ID 7777> which encodes the amino acid sequence <SEQ ID 7778>. Analysis of this protein sequence reveals the following:

```
Possible site: 52

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2993(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC74707 GB:AE000259 glutathionine S-transferase [Escherichia
           coli]
 Identities = 29/137 (21%), Positives = 61/137 (44%), Gaps = 9/137 (6%)

Query: 1   LPFIAKQTLKSQLIPQDNLLAESRFNEIMDFLTGDFPLVFRPMINPHRYTISQDNQALEK 60
           + ++A      QL+   N ++  +  E ++++  +    F P+  P            E+
Sbjct: 70  MQYLADSVPDRQLLAPVNSISRYKTIEWLNYIATELHKGFTPLFRP---------DTPEE 120

Query: 61  VKQASYKRMDIAMTHLDSLIGESGHVYRDQQTIADAYAYAMALWSQKTPKSYENYPHLAA 120
            K     +++  + +++  + +   +   + TIADAY + +  W+     + E   H+AA
Sbjct: 121 YKPTVRAQLEKKLQYVNEALKDEHWICGQRFTIADAYLFTVLRWAYAVKLNLEGLEHIAA 180

Query: 121 FMAKMVEDSAVQQVLNA 137
           FM +M E   VQ  L+A
Sbjct: 181 FMQRMAERPEVQDALSA 197
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2640**

A DNA sequence (GASx932R) was identified in *S.pyogenes* <SEQ ID 7779> which encodes the amino acid sequence <SEQ ID 7780>. Analysis of this protein sequence reveals the following:

```
Possible site: 14

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4081(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2641

A DNA sequence (GASx935) was identified in *S.pyogenes* <SEQ ID 7781> which encodes the amino acid sequence <SEQ ID 7782>. Analysis of this protein sequence reveals the following:

```
Possible site: 45

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.6304(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2642

A DNA sequence (GASx937) was identified in *S.pyogenes* <SEQ ID 7783> which encodes the amino acid sequence <SEQ ID 7784>. Analysis of this protein sequence reveals the following:

```
Possible site: 34

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3503(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2643

A DNA sequence (GASx938R) was identified in *S.pyogenes* <SEQ ID 7785> which encodes the amino acid sequence <SEQ ID 7786>. Analysis of this protein sequence reveals the following:

```
Possible site: 27

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2884(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2644**

A DNA sequence (GASx939) was identified in *S.pyogenes* <SEQ ID 7787> which encodes the amino acid sequence <SEQ ID 7788>. Analysis of this protein sequence reveals the following:

```
Possible site: 50

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2771(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2645**

A DNA sequence (GASx941) was identified in *S.pyogenes* <SEQ ID 7789> which encodes the amino acid sequence <SEQ ID 7790>. Analysis of this protein sequence reveals the following:

```
Possible site: 29

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2257(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2646**

A DNA sequence (GASx942R) was identified in *S.pyogenes* <SEQ ID 7791> which encodes the amino acid sequence <SEQ ID 7792>. Analysis of this protein sequence reveals the following:

```
Possible site: 23

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3255(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```

```
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB91582 GB:AF242881 ymh [Agrobacterium tumefaciens] (ver 2)
 Identities = 75/223 (33%), Positives = 116/223 (51%), Gaps = 2/223 (0%)

Query: 38  DQNSGFNKHKRVHNLVSDILNRTQNTDNIKLVIEYVCNPLRYINEVSIFEQLRTAINIPL 97
            D +    K  R++N  +   N  +   +I   I     P R+  +   FE +R  +N  L
Sbjct: 39  DTDPQMTKRHRLYNAFASDQNSRKQRTHIIAFIRKAMKPERFARDSERFEPMRLNLNRAL 98

Query: 98  SLKGLIVSDSGQIVTTTTSKTLSEAKKRFETLDSRLKELKVHPHVLKFCTQELLQENYFH 157
            +  GL V  SG++     ++TLS+A +R   L + L    VHP VL+FC +ELL +NYFH
Sbjct: 99  AFAGLAVKASGELAAVDAAETLSQATRRALELRADLTSRGVHPDVLRFCREELLVDNYFH 158

Query: 158 AVFEASKGVFHRIRLLTGSAMDSASLIDQCFKPGEPIVIINGNKLQTLDEQSEYKGLKNL 217
            AV EA K V  +IR  TG   D A L+D+ F    P++ I  N+LQ+  E+ E +G  NL
Sbjct: 159 AVLEAVKSVADKIRQRTGLTDDGAVLVDRAFSGDAPMLAI--NELQSESEKGEQRGFSNL 216

Query: 218 LLAIAHLYRNSKAHKLKYYNPDNLNDALTALTLMSLAHNLLDS 260
            +     ++RN+ AH  + +   +  DA    ++ SL H  +D+
Sbjct: 217 VKGTFSMFRNTTAHAPRIHWQMSKEDAEDLFSMFSLMHRRIDA 259
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2647**

A DNA sequence (GASx943R) was identified in *S.pyogenes* <SEQ ID 7793> which encodes the amino acid sequence <SEQ ID 7794>. Analysis of this protein sequence reveals the following:

```
Possible site: 30

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1526(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2648**

A DNA sequence (GASx944) was identified in *S.pyogenes* <SEQ ID 7795> which encodes the amino acid sequence <SEQ ID 7796>. Analysis of this protein sequence reveals the following:

```
Possible site: 19

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1427(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2649**

A DNA sequence (GASx945) was identified in *S.pyogenes* <SEQ ID 7797> which encodes the amino acid sequence <SEQ ID 7798>. Analysis of this protein sequence reveals the following:

```
Possible site: 13

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2578(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC98430 GB:L29324 excisionase [Streptococcus pneumoniae]
 Identities = 23/54 (42%), Positives = 40/54 (73%)

Query: 1  LIQQWEGLTVATAKQWATEMRDHPDFKQFVLNPTHRIVFIDYEGFKLFVQWKSR 54
           ++++W+GL   T  +W  EMR++  F  +V+NPTH++VFI+ EGF+ F++WK +
Sbjct: 21 ILKRWDGLNKYTLNRWIKEMRENRTFSMYVINPTHKLVFINLEGFESFLRWKQK 74
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2650**

A DNA sequence (GASx946) was identified in *S.pyogenes* <SEQ ID 7799> which encodes the amino acid sequence <SEQ ID 7800>. Analysis of this protein sequence reveals the following:

```
Possible site: 16

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -4.99   Transmembrane    3 -  19 (   1 -  23)

----- Final Results -----
               bacterial membrane --- Certainty=0.2996(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2651**

A DNA sequence (GASx950) was identified in *S.pyogenes* <SEQ ID 7801> which encodes the amino acid sequence <SEQ ID 7802>. Analysis of this protein sequence reveals the following:

```
Possible site: 51

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2211(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2652**

A DNA sequence (GASx951) was identified in *S.pyogenes* <SEQ ID 7803> which encodes the amino acid sequence <SEQ ID 7804>. Analysis of this protein sequence reveals the following:

```
Possible site: 30

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4258(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2653**

A DNA sequence (GASx952) was identified in *S.pyogenes* <SEQ ID 7805> which encodes the amino acid sequence <SEQ ID 7806>. Analysis of this protein sequence reveals the following:

```
Possible site: 46

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2476(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF74110 GB:AF212847 ORF245 [Lactococcus lactis bacteriophage
          ul36.2]
 Identities = 82/265 (30%), Positives = 128/265 (47%), Gaps = 27/265 (10%)

Query: 1   MANQLSTQQVKRDITTDPTLLTGADIKKYFDPQNLLSEKQVGQALALCKGRNLNPFANEV 60
            MAN+L    V          L    IK+Y D     S+ ++   + LCK  N+NPF  EV
Sbjct: 1   MANELGIFSVDN--------LNMTTIKQYLDGGGKASDAELVLLINLCKQNNMNPFMKEV 52
```

```
Query: 61  YIVAYKNNSGTDFSLIVSKEAFMKRAERCEGYDGFEAGITVM-RNGEMVEIEGSLKLPDD 119
           Y + Y N       ++VS++ + KRA +   + G E G+ V+ ++G +   EG+ K  +
Sbjct: 53  YFIKYGNQPA---QIVVSRDFYRKRAFQNPNFVGIEVGVIVLNKDGVLEHNEGTFKTHEQ 109

Query: 120 VLIGGWAIVYRKDRSHRYKVTVDFNEYVKLDKYGNPRSTWKSMPGTMIRKTALVQTLREA 179
            L+G WA V+ K+      V V ++EYV++ K G+P   W + P TM+ K A  Q LR A
Sbjct: 110 ELVGAWARVHLKNTEIPVYVAVSYDEYVQM-KDGHPNKMWTNKPCTMLGKVAESQALRMA 168

Query: 180 FPDELGNMYTDIDGGDTFDAIKDVTPQETQEEVRARK---MAQIEQYKQEQ--TQKQTQK 234
           FP E    Y + +  +         P++   EV   K    AQIE + +E    +K  +
Sbjct: 169 FPAEFSGTYGEEEYPE---------PEKEPREVNGVKEPDRAQIESFDKEDYAAKKIEEL 219

Query: 235 ADTSYPVDEVSEHTDDPVQGELLDG 259
            + + P  EV E T + +  E L+G
Sbjct: 220 KEKAQPQKEVVEETGEVIDEEPLEG 244
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2654**

A DNA sequence (GASx953) was identified in *S.pyogenes* <SEQ ID 7807> which encodes the amino acid sequence <SEQ ID 7808>. Analysis of this protein sequence reveals the following:

```
Possible site: 13

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3413(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF74111 GB:AF212847 ORF364 [Lactococcus lactis bacteriophage
           ul36.2]
 Identities = 67/222 (30%), Positives = 120/222 (53%), Gaps = 3/222 (1%)

Query: 1   MQELQLKVTQAQVEIIDREKFEQNINEVVAKYQNYAVTAGTIKDDKQVLADLRKLKKQLS 60
           ++++++    A + I++ EKF+ +IN+VVA+Y  +  +   +  D++  A L KL  ++
Sbjct: 19  VKDIEIDFKPAIINILEEEKFKASINQVVAEYTGHVPSVENLTVDRKTRASLNKLITKIE 78

Query: 61  DERIKVKKELSKPADDIDGYIKQASKPLDDTIDKIATDVKEFEDHQKALRLDTVKSYLSN 120
             R ++KK ++ P  + +G+ K+A  P++  I+ I   +K+ E  QK  R   V   L
Sbjct: 79  TRRKEIKKSINVPYAEFEGWYKKAIAPMEKVIETIDAGIKKIEAEQKESRKKVVHELLVE 138

Query: 121 KASEYMLDPRIFDEKAMEYTKAGNFMADGVTLKKVTMKSLEDLVTFEYQKEQEVEKAKAT 180
             ++  +D RIF+    ++ K+ NF  + +  KK  + S+  ++  E QK  E + AK +
Sbjct: 139 LTTDTEVDSRIFENFVDDWAKSSNF--NDIKPKKQLIDSITYVIDGEKQKIAEYKSAKQS 196

Query: 181 ISGQCAEYGMTDQPYIRMLKE-MTLVEVLGQIKADYLAEKQK 221
           IS  C    +T  PYIRML    T+ E++  I  D L EKQ+
Sbjct: 197 ISDFCFGNNITSTPYIRMLDSGKTVSEIMAVITEDVLFEKQR 238
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2655**

A DNA sequence (GASx954) was identified in *S.pyogenes* <SEQ ID 7809> which encodes the amino acid sequence <SEQ ID 7810>. Analysis of this protein sequence reveals the following:

```
Possible site: 56

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3884(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2656**

A DNA sequence (GASx955) was identified in *S.pyogenes* <SEQ ID 7811> which encodes the amino acid sequence <SEQ ID 7812>. Analysis of this protein sequence reveals the following:

```
Possible site: 34

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.1777(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2657**

A DNA sequence (GASx956) was identified in *S.pyogenes* <SEQ ID 7813> which encodes the amino acid sequence <SEQ ID 7814>. Analysis of this protein sequence reveals the following:

```
Possible site: 16

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.44   Transmembrane   82 -  98 (  81 -  98)

----- Final Results -----
                bacterial membrane --- Certainty=0.1977(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2658**

A DNA sequence (GASx958) was identified in *S.pyogenes* <SEQ ID 7815> which encodes the amino acid sequence <SEQ ID 7816>. Analysis of this protein sequence reveals the following:

```
Possible site: 34

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3673(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2659**

A DNA sequence (GASx960) was identified in *S.pyogenes* <SEQ ID 7817> which encodes the amino acid sequence <SEQ ID 7818>. Analysis of this protein sequence reveals the following:

```
Possible site: 36

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1852(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2660**

A DNA sequence (GASx961) was identified in *S.pyogenes* <SEQ ID 7819> which encodes the amino acid sequence <SEQ ID 7820>. Analysis of this protein sequence reveals the following:

```
Possible site: 45

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.7380(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF63071 GB:AF158600 gp137 [Streptococcus thermophilus
           bacteriophage Sfi11]
 Identities = 67/136 (49%), Positives = 97/136 (71%), Gaps = 2/136 (1%)

Query: 5   PEIDIQKTKSNAKRKLREYPRWRRIANDVDTQKVTATYSFEPRQPHGTPSKPVERLALNR 64
           PEID + T    KRKLREYPRWR IA+D   QK+T  ++F PR   G  +KPVE +A+ R
Sbjct: 4   PEIDEKATLKRCKRKLREYPRWREIAHDSAEQKITQEFTFMPRG--GGVNKPVENIAVRR 61

Query: 65  VSAEQELDTIERAVNGIFDPEYRLILIDKYLLTYPKTDCDIYTKLGYEKSQYYNMLDNAL 124
           V A  EL+ IE+AVNG++ P+YR ILI+KYL   PK +  I   +G+E++ +  +L+N++
Sbjct: 62  VDALNELEAIEQAVNGLYRPDYRRILIEKYLAYPPKPNWQIAQSIGFERTAFQELLNNSI 121

Query: 125 LSFSELYKEGMLLVEK 140
           L+F+ELY++G L+VE+
Sbjct: 122 LAFAELYRDGRLIVER 137
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2661**

A DNA sequence (GASx962) was identified in *S.pyogenes* <SEQ ID 7821> which encodes the amino acid sequence <SEQ ID 7822>. Analysis of this protein sequence reveals the following:

```
Possible site: 16

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3375(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2662**

A DNA sequence (GASx963R) was identified in *S.pyogenes* <SEQ ID 7823> which encodes the amino acid sequence <SEQ ID 7824>. Analysis of this protein sequence reveals the following:

```
Possible site: 48

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2663**

A DNA sequence (GASx964) was identified in *S.pyogenes* <SEQ ID 7825> which encodes the amino acid sequence <SEQ ID 7826>. Analysis of this protein sequence reveals the following:

```
Possible site: 51

>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -6.16   Transmembrane   90 - 106 (  89 - 111)
   INTEGRAL    Likelihood = -5.52   Transmembrane  131 - 147 ( 129 - 150)
   INTEGRAL    Likelihood = -0.43   Transmembrane   53 -  69 (  52 -  69)

----- Final Results -----
               bacterial membrane --- Certainty=0.3463(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has no significant homology with any sequences in the GENPEPT database.

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2664**

A DNA sequence (GASx965) was identified in *S.pyogenes* <SEQ ID 7827> which encodes the amino acid sequence <SEQ ID 7828>. Analysis of this protein sequence reveals the following:

```
Possible site: 15

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3944(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA66779 GB:X98106 Rorf172 [Bacteriophage phig1e]
 Identities = 36/82 (43%), Positives = 52/82 (62%), Gaps = 3/82 (3%)

Query: 18 ELTEKQQRFVDKYITTFNATESAKQAGYSEKSAYSQGQRLLKNVEIQKAMKERFLEAKDT 77
          +LT KQQ+F D+YI + NA ++A++AGYS++SA S GQ  L   +I++ + ER     +
Sbjct: 4  KLTPKQQKFADEYIKSGNAADAARKAGYSKRSARSVGQENLTKPDIKQYIDERM---DEI 60

Query: 78 KGDRIQDVAETLEQDTSIARGE 99
             RI D  E +E  T IARGE
Sbjct: 61 ASKRIMDATEAVELLTRIARGE 82
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2665**

A DNA sequence (GASx966) was identified in *S.pyogenes* <SEQ ID 7829> which encodes the amino acid sequence <SEQ ID 7830>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
```

```
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2389(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB13115 GB:Z99110 PBSX defective prophage terminase (large
           subunit) [Bacillus subtilis]
 Identities = 117/417 (28%), Positives = 195/417 (46%), Gaps = 33/417 (7%)

Query: 31  YRVVKGSRGSKKSKTTALNFIVRLLKYPWANLLVIRRYSNTNKQSTYTDFKWACNQLKVT 90
            Y+ + G  GS KS  TAL  +++LLK      LVIR   +T++ ST+  F+    +L +T
Sbjct: 21  YQFLVGGYGSSKSYHTALKIVLKLLKEK-RTALVIREVFDTHRDSTFALFQEVIEELGLT 79

Query: 91  HLFKFNESLPEITVKATGQKILFRGLDDELKITSITVDVGALCWAWFEEAYQIETEDKFS 150
                   S  ++     G +I+F+G+D+  K+ S    V  +   W EE  +++ E
Sbjct: 80  KAVASLSSPLQLRFH-NGSRIMFKGMDNPAKLKS----VHNISLIWIEECSEVKYEG--- 131

Query: 151 TVVESIRGSLDAPDFFKQITVTFNPWSERHWLKRVFFDEETKR----------------- 193
               + + G L  P+    +  T NP    +W  R FF +E K+
Sbjct: 132 --FKELIGRLRHPELKLHMICTTNPVGTSNWTYRHFFRDERKKRFVLDDSELYEKRTIVK 189

Query: 194 ADTFSGTTTFRVNEWLDDVDKRRYEDLYKTNPRRARIVCDGEWGVAEGLVFDNFEVVDFD 253
             DT+   +T   N +L +   ++ + L + +P   RI   G +GV    V   FEV+  D
Sbjct: 190 GDTYYHHSTANDNLFLPESYVKQLDGLKEYDPDLYRIARKGRFGVNGIRVLPQFEVLPHD 249

Query: 254 -VEKTIQRVKET--SAGMDFGFTQDPTTLICVAVDLANKELWLYNEHYQKAMLTDHIVKM 310
             V+K I  + +     GMDFGF +    ++ +AVD   K L++Y E+YQ  M  D  +
Sbjct: 250 QVKKCIAAISKPIFRTGMDFGFEESYNAVVRLAVDPEKKYLYIYWEYYQNKMTDDRTAEE 309

Query: 311 IRDKNLHRSYIAGDSAEKRLIAEIKSKGVSGIVPSIKGKGSIMQGIQFMQGF-KIYIHPS 369
            +R+    +  I  DSAE + I   + +G   +V + K  GS +Q  + ++ F KI+
Sbjct: 310 LREFIETQELIKADSAEPKSIQYFRQQGFR-MVGARKFPGSRLQYTKKVKRFKKIFCSDR 368

Query: 370 CEHTIEEFNTYTFKQDKEGNWLNEPIDKNNHVIDAIRYALEKYHIRSNESNQFEVLR 426
            CE+ I E  T T+ +DK G  + +    + H + AI YAL+ Y +   +    + +R
Sbjct: 369 CENVIYELETLTYAKDKNGALIEDEFTIDPHTLSAIWYALDDYEVADMKETAHKRMR 425
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2666**

A DNA sequence (GASx967) was identified in *S.pyogenes* <SEQ ID 7831> which encodes the amino acid sequence <SEQ ID 7832>. Analysis of this protein sequence reveals the following:

```
Possible site: 32

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4899(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC34397 GB:AF158600 gp502 [Streptococcus thermophilus
```

```
          bacteriophage Sfi11]
 Identities = 67/114 (58%), Positives = 83/114 (72%)

Query: 6   FRDSTGKTKTLEFRFHREARMRYQAESLESLLTEKYKLLREMIEHHDKVQKPRIQELLDY 65
           F DSTG+   L  RFHRE+R+RY+A++LE L+   ++LL+  I HH   Q PRIQELLDY
Sbjct: 7   FTDSTGQDLVLNLRFHRESRIRYRADNLEELMVNNWELLKNFINHHKLRQAPRIQELLDY 66

Query: 66  AEGNNHTISEIGRRKDDDMADVRAVHNYGKYISTLKQGYLVGNPIRVEYIDGTE 119
           A G NH + + GRRKD++MAD RAVHNYG+ IS  K GYL GNPIRVEY D  +
Sbjct: 67  ARGENHDVLKSGRRKDNEMADKRAVHNYGRMISKFKTGYLAGNPIRVEYDDNED 120
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2667**

A DNA sequence (GASx968) was identified in *S.pyogenes* <SEQ ID 7833> which encodes the amino acid sequence <SEQ ID 7834>. Analysis of this protein sequence reveals the following:

```
Possible site: 34

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4007(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC34397 GB:AF158600 gp502 [Streptococcus thermophilus
          bacteriophage Sfi11]
 Identities = 172/319 (53%), Positives = 227/319 (70%), Gaps = 9/319 (2%)

Query: 1   LIYRSMDDKTEVVRLDPREVFVIYQNNLEQSSLAGVRYYNKNQLDGTTKIVELYTDNKIL 60
           +IYRS  D+T + RL P E FVIY N+LE +S+A VRYYN+  L     +VE+YT+  I
Sbjct: 157 VIYRSEYDETRIKRLSPLETFVIYDNSLEDNSIAAVRYYNRGTLQNAKDVVEIYTNQHIY 216

Query: 61  KFEYDGDLTPIGETSSHAFGSVPITEYLNTDDGMGDYETELSLIDLYDAAQSDTANYMQD 120
             +       I  T  HAFG+VPITE+LN  DG+GDYETEL LIDLYD+A+SDTAN+M D
Sbjct: 217 TLDASDSFNEISVTP-HAFGTVPITEFLNNADGIGDYETELYLIDLYDSAESDTANHMSD 275

Query: 121 LSDAILAIIGRVSFPGYVDTAEKAIEYLRKMRKARLLNLEPPVDQDGREGSVDAKYLYKQ 180
           ++DAILAI G ++ P  +  ++        M++ RL+ L+PP   DG+EG+V A+YL K
Sbjct: 276 MADAILAIYGDLALPQGMQASD--------MKRTRLMQLKPPKSADGKEGTVKAEYLTKS 327

Query: 181 YDVQGTEAYKNRIVSDIHKFTNTPDMTDSKFAGQQSGEALKWKVFGLDQERVDMQALFEQ 240
           YDV G EAYK R+  DIH FTNTPDM+D+ F+G  SGEALK+K+FGLDQ+RVD Q+ F Q
Sbjct: 328 YDVSGAEAYKTRLNKDIHVFTNTPDMSDNHFSGNASGEALKYKLFGLDQDRVDTQSQFTQ 387

Query: 241 SLKRRYKLIARVSQLLKEIDDFDISKLKITFTPNLPKSLQEKIEAFKALGGELSQETAMA 300
            LKRRY+L AR+  L+ E  DFD S+LKITFTPNLPKSL E++     LGG++SQETA++
Sbjct: 388 GLKRRYRLAARIGSLVNEFKDFDESRLKITFTPNLPKSLYEQVSILNDLGGQVSQETALS 447

Query: 301 ITDIVEDAKKEISLINSES 319
           ++ +VE+  +E+  IN ES
Sbjct: 448 LSGLVENPTEELDKINEES 466
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2668**

A DNA sequence (GASx969) was identified in *S.pyogenes* <SEQ ID 7835> which encodes the amino acid sequence <SEQ ID 7836>. Analysis of this protein sequence reveals the following:

```
Possible site: 21

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5307(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC79543 GB:U88974 ORF28 [Streptococcus thermophilus temperate
           bacteriophage 01205]
 Identities = 118/309 (38%), Positives = 183/309 (59%), Gaps = 18/309 (5%)

Query: 8   YWRDRIKKEMDAK-EADDISLEQSMKQLHDYHFRNIEKEIESFYQRYADKEKIDLSEARK 66
            YW  R  +E +A  +  +   ++ ++ L++     + KE++++ Q+YA+K  + +S+A++
Sbjct: 3   YWSKRTLREREASIKKGEAEFKKELEALYNLQLSQLRKELDAYIQKYANKNGLSVSDAKR 62

Query: 67  RASELDISAYQKKAKELVAKAEKLRREGKIVTRDDFTHQENADMSIYNLAMKTNALELLR 126
            +A   D+ A++ KAK  VA               DF+ + N ++  YN +M     ELL
Sbjct: 63  KADSFDVKAFETKAKRYVADK-------------DFSPKANRELQDYNFSMSVGRQELLI 109

Query: 127 LNIDLEMQELANGEHKLTKKFLDEGYRKETEFQAGLLGLSVASQASVKSLADAVINANFK 186
              ++LE+  L+  E +LT  +L  GY+ E   +  LL  +V S  +++    A +NANF+
Sbjct: 110 QELELELLALSESERQLTNDYLTNGYKSEV-VRESLLDQTVPSGKTLEKYMKAAVNANFE 168

Query: 187 GAKWSDNIWDRQDKLRSIISQSVQSAILKGKNGLTIARDIRREFDVSASYAKRLAITEHA 246
            GA+WS+ IW RQ++LR I+   V  A+++G+NGLTIAR IR+  D S + A+RLAITEHA
Sbjct: 169 GAEWSERIWKRQEQLRKIVKTEVTRALIRGENGLTIARRIRKHMDASRTEAERLAITEHA 228

Query: 247 RVQMEVGRLSMAENGFAMFDILPEPKACDVCKDIAKH---GPYHLDKWRIGENSPPFHPY 303
            RVQ       M ENGF  F ++PE +ACD+CKDI K     P  +    IG N+PP HPY
Sbjct: 229 RVQTLAQESIMKENGFEHFKLMPESRACDICKDIGKETEKNPVKIADMEIGTNAPPIHPY 288

Query: 304 CRCAIVGVD 312
            CRCA+V V+
Sbjct: 289 CRCAVVEVE 297
```

Based on this analysis, it was predicted that this GAS-specific protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2669**

A DNA sequence (GASx970) was identified in *S.pyogenes* <SEQ ID 7837> which encodes the amino acid sequence <SEQ ID 7838>. Analysis of this protein sequence reveals the following:

```
Possible site: 15

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2091(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.agalactiae*.